

C	1	48.4	8.1	1685	11	US-11-036-568A-22628	Sequence 22628, A
C	2	46.8	7.8	1092	8	US-10-467-657-4105	Sequence 4105, Ap
C	3	43.2	7.2	1773	11	US-11-036-568A-22025	Sequence 22025, A
C	4	42	7.0	1386	11	US-11-036-568A-17491	Sequence 17491, A
C	5	41.8	7.0	1444	11	US-11-036-568A-26990	Sequence 26990, A
C	6	41.8	7.0	1564	11	US-11-036-568A-20988	Sequence 20988, A
C	7	41.8	7.0	2505	11	US-11-036-568A-19663	Sequence 19663, A
C	8	41.2	6.9	1492	11	US-11-036-568A-10536	Sequence 10536, A
C	9	40.4	6.8	11070	14	US-11-075-185-14	Sequence 34, Appl
C	10	40.4	6.8	78869	14	US-11-075-185-14	Sequence 1, Appl
C	11	40	6.7	1340	11	US-11-036-568A-19408	Sequence 19408, A
C	12	39.6	6.6	1284	9	US-10-714-887-121	Sequence 121, App
C	13	39.4	6.6	1416	8	US-10-858-730-195	Sequence 195, App
C	14	39.4	6.6	1694969	7	US-10-506-454-1590	Sequence 1690, App
C	15	39.2	6.6	908	11	US-11-036-568A-543	Sequence 543, App
C	16	39	6.5	1591	11	US-11-036-568A-19613	Sequence 19613, App
C	17	38.2	6.5	88421	14	US-11-205-109-1	Sequence 1, Appl
C	18	38.2	6.4	1216	8	US-10-821-234-838	Sequence 838, App

C	19	38	6.4	849	11	US-11-0965-5688-21222	Sequence 21222, A
C	20	38	6.4	408	8	US-10-858-730-40	Sequence 40, Appl
C	21	38	6.4	3513	8	US-10-858-730-142	Sequence 142, App
C	22	37.8	6.3	1095	8	US-10-793-626-075	Sequence 3075, Ap
C	23	37.8	6.3	1435	11	US-11-096-5688-11892	Sequence 11892, A
C	24	37.8	6.3	3017	8	US-10-793-626-4422	Sequence 4422, Ap
C	25	37.8	6.3	4185	8	US-10-793-626-5646	Sequence 3646, Ap
C	26	37.8	6.3	5301	14	US-11-075-185-40	Sequence 40, Appl
C	27	37.8	6.3	78659	14	US-11-075-185-1	Sequence 1, Appl
C	28	37.6	6.3	1022	11	US-11-096-5688-10781	Sequence 10781, A
C	29	37.6	6.3	1124	11	US-11-096-5688-23989	Sequence 23989, A
C	30	37.4	6.3	1187	11	US-11-096-5688-22600	Sequence 22600, A
C	31	37.4	6.3	1629	9	US-10-501-035-167	Sequence 167, App
C	32	37.4	6.3	355211	9	US-10-330-773-442	Sequence 242, App
C	33	37.2	6.2	1096	9	US-10-774-887-117	Sequence 117, App
C	34	37.2	6.2	1216	14	US-11-000-688-635	Sequence 635, App
C	35	37.2	6.2	1328	11	US-11-096-5688-8975	Sequence 8975, App
C	36	37.2	6.2	3342	14	US-11-169-041-95	Sequence 95, Appl
C	37	37.2	6.2	157224	14	US-11-112-908-51	Sequence 51, Appl
C	38	37.2	6.2	161726	14	US-11-112-908-48	Sequence 48, Appl
C	39	37.2	6.2	161726	14	US-11-112-908-52	Sequence 52, Appl
C	40	37.2	6.2	170189	14	US-11-112-908-50	Sequence 50, Appl
C	41	37	6.2	765	9	US-10-714-887-125	Sequence 125, App
C	42	37	6.2	1296	11	US-11-096-5688-23788	Sequence 23788, A
C	43	36.8	6.2	1247	11	US-11-096-5688-10198	Sequence 10198, A
C	44	36.8	6.2	1264	11	US-11-096-5688-23146	Sequence 23146, A
C	45	36.8	6.2	2078	14	US-11-024-959-52	Sequence 52, Appl

ALIGNMENTS

```

RESULT 1
US-11-096-568A-22628/c
; Sequence 22628, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theory
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 22628
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1685)
; OTHER INFORMATION: Ceres Seq. ID no. 12409422
US-11-096-568A-22628

```

Query Match	8.1%; Score 48.4; DB 11; Length 1685;
-------------	---------------------------------------

Matches 185; Conservative 0; Mismatches 206; Indels 3; Gaps 1,

Qy	23	GCCCGAGCCGACAGCACTGCTGCAGGATCTCGCCCGATGACAGGTGTGCTTATGAGCCGT	82
Db	933	GCCCGCGCTTGGAGATGCCCTTCGATGACAGCCGCCCGGACCCGCGTGGCCGCGCGCGCGC	874
Qy	83	CGCGGCGCCGACCGCGAGAACTGGCACTGGAGAGGGCGACCCCGCCTGAGAGACGAGACGGAG	142
Db	873	CGGAGCCCGACCGCGGACCCGGCGCACGGCGACAGATGGTGTCTTGTAGAGAGAGAGAGAGG	814
Qy	143	CGGTGTAGCCGTCGTGGCGACCGGATCCGACCCGTCTCGATGAAATTGCGCTCGACCACT	202
Db	813	AGACGCGAGGAGCGGGATCCGGGACGCGCAGCTGCTGAGTAGTGCAGGTACCGGTCC	754
Qy	203	CGAACGAGAACTTTTGTGCTTGTGGCTTCCCTCTGGTGGCACTGCGTCCGCGAGGGTGCAGC	262
Db	753	ACGTCTATGAGCCCTGCGCTCGGCGCCCGGTGTGTGGGGGTCCAGTACACAGTGGACGGCA	694

QY 263 AGAGTGTGAGTGTGCCCCGAGTCTTCAAGCGTGTGGCGAGTCTTTGGCGTCCGAT 322
DB 693 CCACGGGGGGGGCCCCGAGTGTAGCCCAAGCCCCGCCCATATAGGTGAAGTCTTGG 634
QY 323 AAAGCTTGTGAGATGTGCGCTG---CTTGCCTTTCGCGCGCGACGCTCCGCGAGCT 379
DB 633 GAGCTTACACGCTTGTGTGTGTACCGGCGAGCGGAGACCTTGGAACGCGGAGA 574
QY 380 CGAGAAACGACAGGTGGAAGACCGAGCGCGAC 413
DB 573 CGAGCCCCGCGGCTCTGCCCCGTAACCGCGAC 540

RESULT 2

US-10-467-657-4105/c
Sequence 4105, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SPA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIGNANI Vega
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqMln9, version 1.04
SEQ ID NO 4105
LENGTH: 1092
TYPE: DNA
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-4105

Query Match 7.8%; Score 46.8; DB 8; Length 1092;
Best Local Similarity 53.2%; Pred. No. 0.0016;
Matches 99; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 36 CACCTGTGACGATCTGCGCCGCGTAGAGGTGTGCTTGAAGCGCGTCCGCGCGACGC 95
DB 414 CACTGTCCGATGATTTGGCGAGCGGTTAAATGGGTGAAGCCCTGCGCGCGTGA 355
DY 96 GCAGAACTGGCACGGAAAGCGCACCCCGCTGAGACGAGCGAGCGGTGTAGCCGTC 155
DB 354 ACAAATGTACATTCGAAAGCGCACCTTGCGAGGAAATGACAGAGCGTGCAGATC 295
QY 156 GTGGGAGGATCCGACCGCTGTGATGAATTTGCGTTCGACCACTGGAACAGAACTT 215
DB 294 CGATTTCGGGATGAAGAGGTGTTCCAGCGCGTTCCTCCGTACGACATCAAAAGCAATT 235
QY 216 TGTCTG 221
DB 234 TCGAGT 229

RESULT 3

US-11-096-568A-22025
Sequence 22025, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 22025
LENGTH: 1773

TYPE: DNA
ORGANISM: Zea mays subsp. mays
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(1773)
OTHER INFORMATION: Ceres Seq. ID no. 12407294
FEATURE:
NAME/KEY: misc feature
LOCATION: (1742)-(1742)
OTHER INFORMATION: n is a, c, t, g, unknown, or other
FEATURE:
NAME/KEY: misc feature
LOCATION: (1743)-(1743)
OTHER INFORMATION: n is a, c, t, g, unknown, or other
FEATURE:
NAME/KEY: misc feature
LOCATION: (1767)-(1767)
OTHER INFORMATION: n is a, c, t, g, unknown, or other
US-11-096-568A-22025

Query Match 7.2%; Score 43.2; DB 11; Length 1773;
Best Local Similarity 54.0%; Pred. No. 0.017;
Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 241 CTGCGCTGCGGAGGTGACGAGAGGTGTCAGATTGCCGCTGAGTCTTCAAGCTGTG 300
DB 668 CCGCGCGGCGGCGGACCGTGGCGGCGCTCCGCTTCCGCGCTCCGCTCCGCG 727
QY 301 GCCGAGTCTTGGCGCTCCGATTAAGCTTTCGAAAGTGTGCGGTGCTTCGCTGCGCG 360
DB 728 TCCGCGACCTGCGCTGCTTATGCTTCACCTCCGAGAGACGACCTTAAGCGTTGCTG 787
QY 361 CCGACGCGCTCCGCGAGCTCGAGAAACAGAGGTGCAAGAC 401
DB 788 CGAGCGCTTCCGAGACTGTCGCGCGCTGACGACGCGC 828

RESULT 4

US-11-096-568A-17491/c
Sequence 17491, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 17491
LENGTH: 1986
TYPE: DNA
ORGANISM: Zea mays subsp. mays
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(1986)
OTHER INFORMATION: Ceres Seq. ID no. 12358713
US-11-096-568A-17491

Query Match 7.0%; Score 42; DB 11; Length 1986;
Best Local Similarity 46.6%; Pred. No. 0.038;
Matches 135; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 5 CGAGCTTTGACCGCGCGCGCGCGACACCTGTCAGATCTCGCCCGGTAGCA 64
DB 934 CGATGATTTACACGACTTGGCGGAGCGGATGATCTAGTGTGCGATGTGTAGCGGCGG 875
QY 65 GGTGTGCTTGAAGCGCGTGGCGCGCGACGAGAACTGCGACGCGAGCGACCGCG 124
DB 874 CCGCGCGGAGGCGCGCGCGCGACCGCTGACGAAAGGTATGTAGGAGGCGACAG 815
QY 125 CTTGAGACGAGACGAGCGGCTGTAGCCGTGTGCGACGATCCGACCGTCTGATGA 184

Db 814 CGAGAGCGCCCGAAGCCGATCCAGCGGGTACGGTGAACCTTGTCTTCGCGCTTGAGAGA 755
Qy 185 AATTGCGGTGACAGAGTGCAGACAGAACTTTTGTGCTGCGCTTCCCTGCGTGCAGCTGC 244
Db 754 ACTTGTGCGCGGAGAGTTGATCCCAATGCGCGTGAATGAGGGGAGCAAGATTAAGTGGCGA 695
Qy 245 GCTCGGCGAGGGTGCAGAGAGGTGCTGAGTGTCCCGTATGCTTCAAGC 294
Db 694 ACCTGCGGACCTCCCGCGAGAGTCCGCGCTCTGCGCGATGAGACAGAGAGC 645

RESULT 5
US-11-096-568A-26890/c
; Sequence 26890, Application US/11096568A
; Publication No. US20060048240A1

GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26890
; LENGTH: 1147
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1147)
; OTHER INFORMATION: Ceres Seq. ID no. 13628192
US-11-096-568A-26890

Query Match 7.0%; Score 41.8; DB 11; Length 1147;
Best Local Similarity 45.7%; Pred. No. 0.042;
Matches 185; Conservative 0; Mismatches 217; Indels 3; Gaps 1;

Qy 1 AACACAGGCGCTGTGACGCGCGGCGGAGCGCCAGACCTGCTGCAAGATCTGCGCCG 60
Db 676 AACTCATACATAGAGAGAGTGGTGAATGCGACCTGCTGCAAGATGTAATGAACTGC 617
Qy 61 AGCAGGTGTGCTTGAAGCGCTGCGCGCCGACGCGCAGAACTGGACCGGAGGCGCAC 120
Db 616 CGGCGCGCGGAGATGAGAGCGTGCAGAGCGGCGGAGACGTGGGGCGGTGCGAGAG 557
Qy 121 CCCGCTGAGACGAGCGAGCGGCGTGAAGCGCTGCTGCGAGATCCGCACTGCTCG 180
Db 556 AACTGTGCGGATTTGAATGGGCGAGAGCGGCTTGAGCACTGTTGCGATGAGAGGG 497
Qy 181 ATGAATTTGCGCTG--ACCACTGGAACAGAACTTTGTGCTGCGCTTCCCTGCTG 237
Db 496 AGCACTTTGTGCTGCTGACTGAGTCCGAGGAGAGTGAATGTGGGAGGTGCTGAGAG 437
Qy 238 CGACTGCGCTGCGCGAGGCTGCAAGAGGTGCTGAGTTCGCGCTTCAAGTGCAGG 297
Db 436 TCGGGCGGAGAGAGAGAGAGAGCGAGCTGAGATTGACATCTGCAAGTCTTGGTCCCGAG 377
Qy 298 TGGGCGGAGCTTTTGGCTGCGCGATTAAGCTTGTGAGAGTGTGCGCTGCTTCCGCTG 357
Db 376 TTGAGAGAGAGTGTGGCGGCGCGCGGATGTGAGAAATGAAAGGGCTTTCGACAGCAG 317
Qy 358 CGGCGGAGCGCTGCGCGAGCTGCGAGAACTGAGAACTGCAAGCC 402
Db 316 GGCAAGAGAAATGGGTGCTGCGCGGAGTCTCCGAGAGCAC 272

RESULT 6
US-11-096-568A-20988/c
; Sequence 20988, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; TITLE OF INVENTION: Theaby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20988
; LENGTH: 1444
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1444)
; OTHER INFORMATION: Ceres Seq. ID no. 12391561
US-11-096-568A-20988

Query Match 7.0%; Score 41.8; DB 11; Length 1444;
Best Local Similarity 43.5%; Pred. No. 0.042;
Matches 190; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

Qy 15 GAGCGCGGCGCGGAGCGGAGACCTGCTGCAAGATCTGCGCGGTGAAGAGTGTGCTT 74
Db 1107 GAGCGCGGTGAGAGTCCGCGGTGAGTGAATGAACTGAAAGCCAGGTGAGTGCAGAG 1048
Qy 75 GAGCGCGTGCAGCGCGCGGAGCGGAGCTGAGACGCGGAGGCGCACCCGCTGAGAGCA 134
Db 1047 GCTGTGCGGCGCGCGCTTGTCCGTAAGCGGTGATCCGCTGTCCGTGCGTGAATGGGAC 988
Qy 135 GAGCAGAGCGGTGTAAGCGCTGCTGCGAGAGATCCGACCGTCTGATGAATTTGCGCTC 194
Db 987 GCGGTGGCGGAGATGGCGAAGACCGCGTCCAGCTCCACATGACGCTGTGACGGCGT 928
Qy 195 GAGCAGCTGAGACAGAACTTTGTGCTGCGCTTCCCTGCTGCGTCCGCTGCGCGAG 254
Db 927 CATGCGCTGAGAGAGAGCGCGGAGGCGCTTGAAGCGCCAGCGCGCTTCTTGAAGCGC 868
Qy 255 GGTGAGAGAGAGTGTGAGTTCGCGGTGCTTCAAGGTGAGGCGGAGCTTTCG 314
Db 867 GTGAGCTGAGAGAGGCTGCGGCTCCAGAGTGAAGTGAATGTGCTCTCTGCTGCCA 808
Qy 315 CTGCGATTAAGCTTGTGAGAGATGTGCGCTCTTCCGCTTCCGCGGAGCGGCTCCG 374
Db 807 CTCGCGGCGCTGCGGCGCGCGAGCGTGAAGGCGGTGCGGCTGCGCGCGCGCTGCTG 748
Qy 375 GAGCTGAGAGAGAGAGTGTGAGAGAGCGGCGGAGCGGTGCTGCTGCGGAGATGG 434
Db 747 GAGAGAGTGAAGCGCGGAGCGCTGAGAGCGGCTGAGAGTGTGAGAGGCGGCGTGTG 688
Qy 435 TAGACCGAGAGCTGGG 451
Db 687 GTTCCGATGAATGGG 671

RESULT 7
US-11-096-568A-19863/c
; Sequence 19863, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 19863
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(2506)
; OTHER INFORMATION: Ceres Seq. ID no. 12375134
US-11-096-568A-19863

PRIOR APPLICATION NUMBER: US 60/551,103
PRIOR FILING DATE: 2004-03-08
PRIOR APPLICATION NUMBER: US 60/568,290
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.3
SEQ ID NO 1
LENGTH: 78869
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match 6.8%; Score 40.4; DB 14; Length 78869;
Best Local Similarity 44.5%; Pred. No. 0.12;
Matches 161; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 10 CTGTTGACCGCGCGCCCGAGCGGACGCTGTCGACGATCTGCGCCGCTGACGAGTGT 69
DB 12327 CGGCGCGCGCGCGCGCGCGCGAGCGCATCTGTAACCGCTGCGCTGGAACGACGTG 12386
QY 70 CGCTTGAGCGCGCGCGCGCGCGCGAGACGCGGACGCGGAGCGGACCGCGCGCTGA 129
DB 12387 CGCTTGACGACGACGCGCGCTTTGTGCGGAGACGACGTCGTGTCGCGCGGCGCGCG 12446
QY 130 GACGAGACGCGAGCGCGGTGACCGCTGCGGACGCGATCCGACCGCTCTGATGAATTG 189
DB 12447 TGCGCGCGCGCGCTCGGTGACGCGCGGTGCGGCGCGCGCGCGCGCGCTGCGCTGC 12506
QY 190 CGCTGACGACGCTGCGACGAGAACTTTGTGCTGCGCTTCCCTGCTGCGACGTGCGTGC 249
DB 12507 CGGACGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGACGCTGACGCGCGGCGCG 12566
QY 250 GCGAGGAGTGCAGCAGAGGTGCTGAGTTGCCGTAAGCTTTCAGCGGTGCGCGGAGTCT 309
DB 12567 CGCTGCTGACGCGCGCTTGCACGCGCGCGCGCGCGCGCTGCTGCTGCGAGGCGTGC 12626
QY 310 TTGCGCTGCGCGATGAAGCTTTGCGAGATGTCGCTGCTGCGCTTCCGCGCGCGCGCG 369
DB 12627 TGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTCTGCTGCTGACGCGCGCGCGCGTGC 12686
QY 370 TC 371
DB 12687 TC 12688

RESULT 11
US-11-096-568A-19408/C
Sequence 19408, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
PRIOR FILING DATE: 2005-04-11
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 19408
LENGTH: 1340
TYPE: DNA
ORGANISM: Zea mays subsp. mays
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(1340)
OTHER INFORMATION: Ceres Seq. ID no. 12373073
US-11-096-568A-19408

Query Match 6.7%; Score 40; DB 11; Length 1340;
Best Local Similarity 46.4%; Pred. No. 0.14;
Matches 150; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 3 CACGAGCTGTTGACGCGCGCGCGCGCGCGCGACGCTGCGACGATCTCGCGCGGTAG 62

DB 1089 CACCGCCCAATTATGACGAGGAGCTTAAGCAGAGCGGCGCCACGCTCAGCGCTGAGGA 1030
QY 63 CAGGTTCGCTTGAAGCCGCTCGCGCCCGACGCGGACGAACTGGCAGCGGAGCGCACCC 122
DB 1029 GCTACTAATGACACGCTTTGTTGTGTTGCTGCGATGAAGTACACCGCGCGCGCC 970
QY 123 CGCTGAGACGAGACCGAGCGGTGAGCGCTGTCGTCGCGACGCGATCGGACCTTCGAT 182
DB 969 CGACGAGCGAGCGGCTGCGCTGCTGATCGCTGACGACCCCAAGCTCTCTGCGCGTGC 910
QY 183 GAAATTCGCTGACGACGCTGGAACGAACTTTGCTGTCGCTTCCCTGTCGACT 242
DB 909 GTACATGCGCGCGCGGAACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
QY 243 GCGCTCGCGGAGGTGACGAGAGGCTGTCGACTTCCCG 282
DB 849 GGGGCGCGCGCTGACGAGCGGAGCGCGGTGAGATGCCG 810

RESULT 12
US-10-714-887-121/C
Sequence 121, Application US/10714887
Publication No. US20060015972A1
GENERAL INFORMATION:
APPLICANT: Mendel Biotechnology, Inc.
APPLICANT: HEARD, Jacqueline
APPLICANT: RIECHMANN, Josef Luis
APPLICANT: CREELMAN, Robert
APPLICANT: RATCLIFFE, Oliver
APPLICANT: CANALES, Roger
APPLICANT: REPERTI, Peter
APPLICANT: KUMIMOTO, Roderick W
APPLICANT: GUTTERSON, Neal
APPLICANT: REUBER, T. Lynne
APPLICANT: PINEDA, Omaira
APPLICANT: SHERMAN, Bradley K
TITLE OF INVENTION: PLANT TRANSCRIPTIONAL REGULATORS OF DROUGHT STRESS
FILE REFERENCE: MB10058-CIP
CURRENT APPLICATION NUMBER: US/10/714,887
PRIOR FILING DATE: 2003-11-13
PRIOR APPLICATION NUMBER: 10/412,699
PRIOR FILING DATE: 2003-04-10
PRIOR APPLICATION NUMBER: 09/506,720
PRIOR FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/135,134
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: 09/394,519
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: 09/533,392
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,029
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/532,591
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,030
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/125,814
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 09/713,994
PRIOR FILING DATE: 2000-11-16
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 430
SOFTWARE: PatentIn version 3.2
SEQ ID NO 121
LENGTH: 1284
TYPE: DNA
ORGANISM: Oryza sativa (japonica cultivar-group)
FEATURE:
OTHER INFORMATION: G3690 Predicted polypeptide sequence is orthologous to G2999
US-10-714-887-121

Query Match 6.6%; Score 39.6; DB 9; Length 1284;
Best Local Similarity 49.1%; Pred. No. 0.18;

Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 18 GCGGCGCCGAGCGCCGACACCTGTCGACGATGCGCCGCGTAGACAGTGTGCTTGG 77
DB 675 GCGGCGCCGAGCGCCGACGCGCGCGCGACGACCGCGCTCGCGCGGTGGAAGCT 616
QY 78 GCGGCGCCGAGCGCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 137
DB 615 ACAGTGGCAGCG 556
QY 138 GCAAGCGCGGTAGCG 197
DB 555 GAACTCCCG 496
QY 198 CAGCTGAGAGAGAACTTGTGCTGCTGCTTCC 231
DB 495 CTCCCTGTACTCCAGCTCG 462

RESULT 13

US-10-858-730-195/C
Sequence 195, Application US/10858730
Publication No. US20050255568A1
GENERAL INFORMATION:
APPLICANT: Bailey, Richard B.
APPLICANT: Blomquist, Paul
APPLICANT: Doten, Reed
APPLICANT: Driggers, Edward M.
APPLICANT: Madden, Kevin T.
APPLICANT: O'Leary, Jessica
APPLICANT: O'Toole, George
APPLICANT: Trueheart, Joshua
APPLICANT: Walbridge, Michael J.
APPLICANT: Yorgey, Peter S.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
FILE REFERENCE: 14184-030001
CURRENT APPLICATION NUMBER: US/10/858,730
CURRENT FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/475,000
PRIOR FILING DATE: 2003-05-30
PRIOR APPLICATION NUMBER: US 60/551,860
PRIOR FILING DATE: 2004-03-10
NUMBER OF SEQ ID NOS: 364
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 195
LENGTH: 1416
TYPE: DNA
ORGANISM: Thermobifida fusca
US-10-858-730-195

Query Match 6.6%; Score 39.4; DB 8; Length 1416;
Best Local Similarity 47.4%; Pred. No. 0.2;
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 165 GATCGGACCGCTCTGATGAATGCGCTGACACGCTCGAAGGAACTTGTGCTGTG 224
DB 1134 GGTCAACGCGCGGTGGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
QY 225 GCTTCCCTGGTGGACATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTA 284
DB 1074 ACCG 1015
QY 285 GTGCTTACGCGGTGGCG 344
DB 1014 TTGACACGACGAGTATTCACGCGTCAACGCGTCAAGGTGCGGCGCGGTGATGATACCGC 955
QY 345 TGCTTGGCGGTGGCG 404
DB 954 GCGGACATGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
QY 405 CGGCGCGAC 413

DB 894 GGCACCGCG 886

RESULT 14

US-10-506-454-1690/C
Sequence 1690, Application US/10506454
Publication No. US20060068386A1
GENERAL INFORMATION:
APPLICANT: Slesarev, Alexi I
APPLICANT: Mezhevaya, Katia V
APPLICANT: Polushin, Nikolai N
APPLICANT: Shcherbina, Olga V
APPLICANT: Shakhova, Vera V
APPLICANT: Malykh, Andrei A
APPLICANT: Kozavkin, Sergei A
TITLE OF INVENTION: The Complete Genome and Protein Sequences of the Hyperthermophil
TITLE OF INVENTION: Methanopyrus Kandleri AV19 and Monophyly of Archaeal Methanogens
TITLE OF INVENTION: and Methods of Use Thereof
FILE REFERENCE: FID001
CURRENT APPLICATION NUMBER: US/10/506,454
CURRENT FILING DATE: 2004-08-31
PRIOR APPLICATION NUMBER: PCT/US03/06664
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: 60/361,742
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 1722
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1690
LENGTH: 1694969
TYPE: DNA
ORGANISM: Methanopyrus kandleri
US-10-506-454-1690

Query Match 6.6%; Score 39.4; DB 7; Length 1694969;
Best Local Similarity 57.9%; Pred. No. 0.25; Indels 51; Gaps 0;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 282 GTAGTGTTCAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 341
DB 11340 GTGAGCGCGCATCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11281
QY 342 GCGTGTGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401
DB 11280 GACCACCGGAACTCCACGCTCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 11221
QY 402 C 402
DB 11220 C 11220

RESULT 15

US-11-096-568A-5433/C
Sequence 5433, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 5433
LENGTH: 908
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(908)
OTHER INFORMATION: Ceres Seq. ID no. 14309488
US-11-096-568A-5433

Query Match 6.6%; Score 39.2; DB 11; Length 908;

Best Local Similarity 54.1%; Pred. No. 0.23;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY      78  GCCGTGCGGGCCGACGCGAGAACTGGACGCGAAGGCCGACCCCGCTGAGACGAGAC 137
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      531  GGGGACGAGCACCOCGCGACGACGCGGCAAGAGATGCGAGCGACGCGGCTCGAC 472
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      138  GCAGCGGCTGTAGCCGTCGTGGCGAGGATCCGACCGCTCGATGAATTTGCCGCTGAC 197
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      471  GCAGGCCGTGTGAAACGCGTGGCTGCACTGCGGCGACCCGCAATGCTGCGCTCTC 412
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      198  CAGCTCGAACGAGAACTTTGTCGCTTG 225
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      411  GAACTCGGCGAGGAGATGGCGCACTCG 384
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Search completed: April 7, 2006, 03:50:18
Job time : 229.02 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 2299.02 seconds
(without alignments)
12149.485 Million cell updates/sec

Title: US-10-611-442-2_COPY_17000_17596

Perfect score: 597

Sequence: 1 aacacgagcctgtgacgc.....cacgaacgcgcagcgcgcg 597

Scoring table: IDENTITY NUC

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61.2	10.3	188	7	CN026633 UMC-p4civ
2	56.4	9.4	773	10	CL685281 PRI01404
3	56	9.4	883	7	CK151489 FGAS03405
4	55.2	9.2	561	9	BH899909 OCS00976
5	54.4	9.1	872	7	CK152875 FGAS03595
6	54.4	9.1	925	10	CNS00091P
7	54.4	9.1	1133	8	DR741095 FGAS00102
8	54.2	9.1	717	3	BI955767 HVSMM002
9	53.8	9.0	655	8	CV948578 PVP0112
10	53.8	9.0	671	8	CV948570 PVP0112
11	53.4	8.9	849	7	BR265573 HV CEA001
12	53.4	8.8	857	7	CK152637 FGAS03571
13	52.8	8.8	339	2	BE415712 MML038.G0
14	52.8	8.8	528	6	CD922535 G750.1037
15	52.8	8.8	771	7	CK124607 BBS182410
16	52.8	8.7	537	3	BM322439 P1C1_4.G1
17	52	8.7	532	6	CF432914 N1T1_19.F
18	52	8.7	541	1	AM924660 NS1_71.H0
19	52	8.7	608	7	CN151959 WOUNDI_78
20	52	8.7	622	6	CF432927 N1T1_19.G
21	52	8.7	635	7	CN150225 WOUNDI_67
22	52	8.7	705	6	CF429132 PH1_19.H0

C 23	52	8.7	712	7	CN142245	CN142245 WOUNDI_4
C 24	52	8.7	743	7	CN140012	CN140012 OX1_33.B0
C 25	52	8.7	765	3	CD432831	CD432831 ETH1_33.G
C 26	51.4	8.6	925	10	CNS00091P	AL053013 Drosophila
C 27	51	8.5	351	7	CN026960	CN026960 UMC-p4civ
C 28	51	8.5	357	7	CN027260	CN027260 UMC-p4civ
C 29	50.4	8.4	648	6	CD233094	CD233094 SSI_11.H1
C 30	50.4	8.4	657	6	CD233402	CD233402 SSI_13.CO
C 31	50.4	8.4	663	6	CD233416	CD233416 SSI_13.BO
C 32	50.2	8.4	627	7	CA085330	CA085330 SCJLAW209
C 33	49.4	8.3	348	7	CN026914	CN026914 UMC-p4civ
C 34	48.4	8.1	642	10	CG381651	CG381651 OGWD787H
C 35	48.4	8.1	710	9	BZ529592	BZ529592 OGJUR36TM
C 36	48.4	8.1	720	9	BZ529589	BZ529589 OGJUR36TC
C 37	48.4	8.1	794	9	CC668550	CC668550 OGJUR36TV
C 38	48.2	8.1	581	6	CB213001	CB213001 OMT03281
C 39	47.2	7.9	669	8	DR800406	DR800406 ZM_BFB002
C 40	47.2	7.9	677	8	DR807853	DR807853 ZM_BFB003
C 41	47.2	7.9	680	6	CA195481	CA195481 SCE2S109
C 42	47.2	7.9	681	9	CC678957	CC678957 OGHG922TH
C 43	47.2	7.9	771	7	CO093637	CO093637 GR_BA15F
C 44	47.2	7.9	862	2	BG310173	BG310173 HVSME001
C 45	47.2	7.9	870	9	CC637331	CC637331 OGCX32TH

ALIGNMENTS

RESULT 1
LOCUS CN026633 188 bp mRNA linear EST 09-SEP-2004
DEFINITION UMC-p4civ1-006-e01 4-Cell-Embryo-(in vivo) p4civ Sus scrofa CDNA
3', mRNA sequence.

ACCESSION CN026633
VERSION CN026633.1 GI:48721169
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 188)
AUTHORS Whitworth,K., Springer,G.K., Forrester,L.J., Spollen,W.G., Ries,J., Lamberson,W.R., Blvens,N., Murphy,C.N., Mathaiagan,N., Green,J.A. and Prather,R.S.

TITLE Developmental expression of 2489 gene clusters during pig embryogenesis: an expressed sequence tag project
JOURNAL Biol. Reprod. 71 (4), 1230-1243 (2004)
PUBMED 15175238

COMMENT Contact: DNA Core Facility (Swine Project)

Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: porcine@net.missouri.edu
POLYA=yes.

FEATURES
source Location/Qualifiers

1..188
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone_lib="p4civ"

/note="Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Slaughterhouse-derived oocytes were collected, and after cumulus cell removal were used for germinal vesicle stage oocytes, or were matured in vitro (with cumulus cells attached), in vitro fertilized and cultured. In vivo

produced 4-cell and blastocyst stage embryos were collected on days 3 and 6, respectively. Zoonc pelliculide were removed from the embryos prior to mRNA isolation. Expanded descriptions of how the tissues were collected

(default parameters) has been run on this sequence. Lucy identified the region [100,563].
Plate: Talt316 row: e column: 09.
Location/Qualifiers

FEATURES

source

1..883
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultiivar="Wheat line CI 14106"
/db_xref="taxon:4565"
/lab_host="DHS alpha"
/clone_1lb="Triticum aestivum FGAS: Talt3"
/note="Organ: Crown; Vector: pGEN-T; SSH (suppression subtractive hybridization) cDNA library from genotype CI14106 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H) (driver). Nitro-pyrole anchored oligo-dT priming and non-directional cloning."

ORIGIN

Query Match 9.4%; Score 56; DB 7; Length 883;
Best Local Similarity 58.6%; Pred. No. 0.0015; Indels 12; Gaps 1;
Matches 123; Conservative 0; Mismatches 75;

QY 24 CCCGAGCGCCGACGACCTGTGACGATCTGCCCGGTAGAGGTTGCTTGAGGCCGTC 83
DB 427 CTCTATGGCCAAACCTGCTCGACATCTGTGCCCTTAAAGTTCTCCGAGCCTCC 368
QY 84 GCGGCCGACGCGCAAACTGSCACCGGAGGCGCACCCCGCTGAGACGACGACGCG 143
DB 367 CTTCAGTGGGACAGACGACGACGAGGGGAGCGACGACCTGTGACGAAACACAGGC 308
QY 144 GGTGTAGCC-----GTGCTGGGACGACGATTCGACCGCTCTCGATGAAATTGCC 191
DB 307 GGTGACCTTAGGGGTGCCCTGTGTCGCAACGAGGATTCGCCACGCTCCACCAACTGTT 248
QY 192 GTGACCAAGCTCGAACGACGAACTTTGTCGT 221
DB 247 GTCTCTCACTTGAGAAAGTATCTTGTGTGT 218

RESULT 4 561 bp DNA linear GSS 30-AUG-2002
BH899909/c
LOCUS Ots800976 Ostreococcus tauri genomic shotgun library Ostreococcus
DEFINITION tauri genomic clone Ots800976, 5', genomic survey sequence.
ACCESSION BH899909
VERSION BH899909.1 GI:22551404
KEYWORDS GSS.
SOURCE Ostreococcus tauri
ORGANISM Ostreococcus tauri
Bukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Mamiellales;
Mamiellaceae; Ostreococcus.
1 (bases 1 to 561)
Derelle, E., Ferraz, C., Lagoda, P., Eychenne, S., Cooke, R., Regad, F.,
Saban, X., Courties, C., Delseny, M., Demaille, J., Picard, A. and
Moreau, H.
DNA libraries for sequencing the genome of Ostreococcus tauri
(Chlorophytae, Prasinophyceae): the smallest free-living eukaryotic
cell
J. Phycol. 38 (6), 1150-1156 (2002)
Contact: Moreau H
Laboratoire Arago
CNRS UMR 7628
BP 44, Avenue Fontaule, 66651 Banyuls sur mer, France
Tel: (33) 468887309
Fax: (33) 468887398
Email: h.moreau@obs-banyuls.fr
Seq primer: reverse
Class: shotgun.

FEATURES

source

Location/Qualifiers
1..561
/organism="Ostreococcus tauri"

ORIGIN

Query Match 9.2%; Score 55.2; DB 9; Length 561;
Best Local Similarity 65.3%; Pred. No. 0.0024; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 43;

QY 24 CCCGAGCGCCGACGACCTGTGACGATCTGCCCGGTAGAGGTTGCTTGAGGCCGTC 83
DB 529 CTCAACTCTAGACCTGCGACGACGATCTCACTGCCGAGAGGTTCCGCTTGAAGCCTTG 470
QY 84 GCGGCCGACGCGCAAACTGSCACCGGAGGCGCACCCCGCTGAGACGACGACGCG 143
DB 469 CACTTCAGTGGGACAGAAACGAACTTCACTGCGGATCTTACCTGCGATGACACAGCG 410
QY 144 GGTG 147
DB 409 CGTG 406

RESULT 5 872 bp mRNA linear EST 05-DEC-2003
CK152875/c
LOCUS FGAS035952 Triticum aestivum FGAS: Talt3 Triticum aestivum cDNA,
DEFINITION mRNA sequence.
ACCESSION CK152875
VERSION CK152875.1 GI:38972365
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 872)
Allard, F., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,
Genswein, B., Grati, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Link, M.G., McCarthy, E.L., Monroy, A., Murak, I., Nilson, D.,
Peniket, C., Roach, J.L. and Sahran, F.
Functional Genomics of Abiotic Stress In Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: Fgaa-ests@cs.usask.ca

This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [124,739].
Plate: Talt351 row: A column: 23.
Location/Qualifiers

FEATURES

source

1..872
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultiivar="Wheat line CI 14106"
/db_xref="taxon:4565"
/lab_host="DHS alpha"
/clone_1lb="Triticum aestivum FGAS: Talt3"
/note="Organ: Crown; Vector: pGEN-T; SSH (suppression subtractive hybridization) cDNA library from genotype

ORIGIN

CI14106 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, heated and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H) (driver). Nitro-pyrole anchored oligo-dT priming and non-directional cloning."

Query Match 9.1%; Score 54.4; DB 7; Length 872;

Best Local Similarity 58.1%; Pred. No. 0.0039;

Matches 122; Conservative 0; Mismatches 76; Indels 12; Gaps 1;

24 CCGAGCGCCAGCCTGCTGACGATCTGCCCCGCTAGCAGTGTGCTTGAAGCCGTC 83
451 CTCTATGGCCAAAGACTGCTGCAATCTGCCCCCTTAGAGTTCTCCGGAACCTCTCC 392
84 GCGGCGCCAGCGGAGAACTGTCACCGGACCCCGCTGAGACGAGACGAGCC 143
391 CTCCCAATGTCACGAAACGAGCAACGAGGAGCGGACCTGTACGAAACAGAGC 332
144 GGTGTAGCC-----GTGTCGACGAGATCCGACCGTCTGATGAATTTGCC 191
331 GGTAGAGCTTACGCGCGCCCTGTCTCAACGGGATCCCAACGCTTCCACCACTGTT 272
192 GTGACCGAGCTGAAACAGAACTTTGTCT 221
271 GTCCCTCCAACTTGAGAGTATCTTGTTGT 242

RESULT 6
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR1916 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AT053013
VERSION AT053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLES Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT

- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source location/Qualifiers

1..925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR1916"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 9.1%; Score 54.4; DB 10; Length 925;
Best Local Similarity 11.4%; Pred. No. 0.004;
Matches 42; Conservative 180; Mismatches 148; Indels 0; Gaps 0;

9 CTTGTGACCCCGGCGCCCGACACCTGCTGACGATCTCCCGCGTAGAGTG 68
555 SSGSGYKGGSSGSGBSGCCSSSCSSSCCBCCCCGCSYCCSSBSBSKST 614
69 TCGGTAGGCGCCGCGCCCGACGAGAACTGTCACGAGGAGCGGACCGGCTG 128
615 BSGSCSSSKSVCTGTCSSSSSCSSSSSTSSSTSTSSKSSSGSSSSSYTTSK 674
129 AGACGAGACCGAGCGGTGTAGCCGCTGTGACGAGATCCGACCGTCTGATGAATT 188
675 TSAAGSGSWAGGSGSTGTSSTSSSVSSGSKSTBSGGBSSSGSSSS 734
189 GCGGTGACGACGCTGCAACGAACTTTGTGCTGCTTCCCTGATGAGCTGCGCTC 248
735 STSBSBSTSTSSSSSVSSSTGCTCCCSYSSSTSSSTSSSTSSSTSSSVGT 794
249 GCGAGGAGTCGACGAGATGCTGCAAGTCCCGCTAGCTTACCGCTGTGAGCGGATC 308
795 SSSSDSTGSCCCCTTCCTCCTTBMBCYTSTGCGSSSSGKGVTKCGCGGSSSTN 854
309 TTTCGCTGCGGATTAAGCTTGTGAAAGATGTCGCTTCCGCTTCCCGCGGACCG 368
855 GMBGTSSACSSSSSSSSSVSSSKSSSVSSSGSVSSSSASKSSSGSVSS 914
369 CTCGCGAGC 378
915 GSGSGSGSVS 924

RESULT 7

DR741095/c 1133 bp mRNA linear EST 18-JUN-2005
LOCUS FGAS001026 Triticum aestivum FGAS: Library 2 Gate 3 Triticum
DEFINITION aestivum cDNA, mRNA sequence.

ACCESSION DR741095
VERSION DR741095.1 GI:70970528
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poidea; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 1133)
AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
Peniket, C., Roach, J.L. and Sarhan, F.

TITLE Functional Genomes of Abiotic Stress in Wheat and Canola Crops
JOURNAL Unpublished (2003)
CONTACT: Patrick Gulick
Plant Molecular Biology
Concordia University, Department of Biology
7141 Sherbrooke St. West, Montreal, Quebec H4B 1A6, Canada
Tel: 514 848 2424 Ext 3407
Fax: 514 848 2881
Email: pgulick@alcor.concordia.ca

This sequence is the direct result of the Base calling software phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [32,747].
Plate: L2B0204 row: J column: 19.

FEATURES
Source location/Qualifiers

1..1133
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 2 Gate 3"
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Aerial

parts (crown and leaf) of wheat cultivar Norstar from control and long exposure times to low temperature. 4 mRNA populations were combined before constructing the library; 7 days non-acclimated plants and 1, 23, and 53 days cold-acclimated at 4°C. Non-acclimated and cold-acclimated plants were grown in vermiculite. This is the only library that was done according to the Invitrogen manual, and therefore, a percentage of clones will not have the 3 prime end because of NotI digestion within the cDNA."

ORIGIN

Query Match 9.1%; Score 54.4; DB 8; Length 1133;
Best Local Similarity 58.1%; Pred. No. 0.004;
Matches 122; Conservative 0; Mismatches 76; Indels 12; Gaps 1;

QY 24 CCGAGGCGCAGACCTGCTGTCAGATCTGCCCGGTAGACAGTGTGCTTGAAGCGCTC 83
DB CTTATATGCGCAATACCTGCTGACATCTGTCCTTGAAGTTCCTCCAAACCTCC 527
QY 84 GCGGCGCGACGCGAGACTGCGACCGGAGGCGCACCGCGCTGAGACGACGCGAGC 143
DB 526 CTTCCAGTGGCAGACGACGACGACGAGGCGGACCGCAACCTGCATGAAAGCGAGC 467
QY 144 GGTGTAGCC-----GTCTGGCGACGATCCGACGCTCTGATGAATTGCC 191
DB 466 TGTGAGTCTAGGGGTGCGCTGCTGTCAGAGGGGATCCCAACGCTCCACAGTCTTT 407
QY 192 GTGACCAAGCTCGACAGGAACTTTGTCT 221
DB 406 GTCTCCAGCTGAGAAATCTTGTGTGT 377

ORIGIN

/issue type="green seedling leaf"
/lab host="TJ121"
/clone lib="Hordeum vulgare green seedling EST library
HVCDA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TV Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excluded to give pluscript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Fritsch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TV, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30
(<http://wheat.pw.usda.gov/gspages/bgn/31/cover.html>)"

RESULT 8
BI955767/c 717 bp mRNA linear EST 19-OCT-2001
LOCUS HVSMEM0024G19f Hordeum vulgare green seedling EST library
DEFINITION HVCDA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
clone HVSMEM0024G19f, mRNA sequence.
BI955767
BI955767.1 GI:16302370
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 717)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Fritsch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
Simmons,J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
Unpublished (2001)
Contact: Wing R.A
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 310
Seg primer: AATTAACTCTCACTAAAGG
High quality sequence start: 4
High quality sequence stop: 551.
Location/Qualifiers

FEATURES

1..717
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEM0024G19f"

Query Match 9.1%; Score 54.2; DB 3; Length 717;
Best Local Similarity 58.5%; Pred. No. 0.004;
Matches 120; Conservative 0; Mismatches 73; Indels 12; Gaps 1;

QY 29 GCGCCAGACCTGCTGTCAGATTCGCCCGGTAGACAGTGTGCTTGAAGCGCTCGCGC 88
DB 557 GCGCCAGACCTGCTGTCAGATTCGCCCGGTAGACAGTGTGCTTGAAGCGCTCGCGC 498
QY 89 CCGAGCGGAGAACTGCGACGCGAGCGGCGCGCTGAGACGAGCGAGCGGTGT 148
DB 497 CTGTGCAAAACGACGACGAGGAGGCGCGCTGTGTGTAAGCAGCGGTGA 438
QY 149 AGCC-----GTCTGGCGACGATCCGACGCTCTGATGAATTCCTCGCA 196
DB 437 GCTTAGCGGTGCCCTGTGTGTGACAGGATCCCAACGCTCTTGTCTT 378
QY 197 CCACTCGAAGCAAGAACTTTGTCT 221
DB 377 CCAGCTTGAGAAATATCTTGTGTGT 353

RESULT 9
CV948578/c 655 bp mRNA linear EST 25-JAN-2005
LOCUS PVrpbv 12064 zoospores, purified Phytophthora infestans cDNA, mRNA
DEFINITION
ACCESSION CV948578
VERSION CV948578.1 GI:58138334
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 655)
Randall,T., Dwyer,R.A., Hultema,B., Beyer,K., Cytanich,C.,
Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,B., Gaffney,T.,
Law,M., Teale,A., Torto-Alalibo,A., Zhang,M., Zheng,L., Mueller,E.,
Windass,J., Binder,A., Birch,P.R.J., Gisi,U., Govers,F., Gow,N.A.,
Mauch,F., Van West,P., Maugh,M.E., Yu,J., Bolter,T., Kamoun,S.,
Lam,S.T. and Judelson, H.S.

TITLE Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi

JOURNAL Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)

PUBMED 15782637

COMMENT Contact: Judelson HS
Department of Plant Pathology
University of California
Webster Hall, Riverside, CA 92521, USA
Tel: 909 787 4199
Fax: 909 787 4294
Email: howard.judelson@ucr.edu.

FEATURES Location/Qualifiers

source 1..655
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="88069"
/db_xref="taxon:4787"
/sex="M"
/clone_lib="zoospores, purified"
/note="Vector: pSPORI1"

ORIGIN

Query Match 9.0%; Score 53.8; DB 8; Length 655;
Best Local Similarity 55.7%; Pred. No. 0.0055;
Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 37 ACTGCTGCAGATCTCCGCCGCTGACAGGTGTCCTTGAGCCGCGCCGACGCC 96
DB 491 AGCTGCTCCAGATCTCCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 432
QY 97 CAGACTGCGACGCGGACGCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 156
DB 431 CAGAAATGCGACGCGGACGCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 372
QY 157 TGGCAGCGATCCGCGACGCGCTGATGAATTCGCTGACGCGCTGACGAGAACTTT 216
DB 371 GTGTGGCGATGATCAAGCGCTTGCATTCGTCCTCATTTGAAGTTTAAGCAGTAATTC 312
QY 217 GTCGT 221
DB 311 ACCGT 307

RESULT 10 CV948570 671 bp mRNA linear EST 25-JAN-2005
LOCUS PVXpvb.12056 zoospores, purified *Phytophthora infestans* cDNA, mRNA
DEFINITION sequence.
ACCESSION CV948570 GI:58138326
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE 1 (bases 1 to 671)
Randal, T., Dwyer, R.A., Huitema, E., Beyer, K., Cvitanich, C., Kelkar, H., Fong, A.M., Gates, K., Roberts, S., Yacskan, E., Gaffney, T., Law, M., Testa, A., Torto-Alalibo, A., Zhang, M., Zheng, L., Mueller, E., Windass, J., Binder, A., Birch, P.R.J., Gisi, U., Govers, F., Gow, N.A., Mauch, F., van West, P., Waugh, M.E., Yu, J., Bolter, T., Kamoun, S., Lam, S.T., and Judelson, H.S.
Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi

TITLE Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)

JOURNAL 15782637

PUBMED Contact: Judelson HS

COMMENT Department of Plant Pathology
University of California
Webster Hall, Riverside, CA 92521, USA
Tel: 909 787 4199

FEATURES Location/Qualifiers

source 1..671
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="88069"
/db_xref="taxon:4787"
/sex="M"
/clone_lib="zoospores, purified"
/note="Vector: pSPORI1"

ORIGIN

Query Match 9.0%; Score 53.8; DB 8; Length 671;
Best Local Similarity 55.7%; Pred. No. 0.0055;
Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 37 ACTGCTGCAGATCTCCGCCGCTGACAGGTGTCCTTGAGCCGCGCGCCGACGCC 96
DB 448 AGCTGCTCCAGATCTCCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 389
QY 97 CAGACTGCGACGCGGACGCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 156
DB 388 CAGAAATGCGACGCGGACGCGCCGCTGACAGGTGTCCTTGAGCCGCGCGACGTGCG 329
QY 157 TGGCAGCGATCCGCGACGCGCTGATGAATTCGCTGACGCGCTGACGAGAACTTT 216
DB 328 GTGTGGCGATGATCAAGCGCTTGCATTCGTCCTCATTTGAAGTTTAAGCAGTAATTC 269
QY 217 GTCGT 221
DB 268 ACCGT 264

RESULT 11 BF265573 849 bp mRNA linear EST 23-OCT-2001
LOCUS HV_CEA0012L01f Hordeum vulgare seedling green leaf EST library
DEFINITION HVCDNA0004 (Blumeria challenged) Hordeum vulgare subsp. vulgare cDNA clone HV_CEA0012L01f, mRNA sequence.
ACCESSION BF265573
KEYWORDS BF265573.1 GI:11196567
SOURCE EST.
ORGANISM Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 849)
Wing, R., Close, T.J., Kleinhofe, A., Wise, R., Wei, F., Begum, D., Fritsch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Oates, R., and Main, D.
Development of a genetically and physically anchored EST resource for barley genomes: Blumeria infected incompatible (ML13) seedling leaf cDNA library
Unpublished (2001)

TITLE Contact: Wing RA
JOURNAL Clemson University Genomics Institute
COMMENT Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 448
Seq primer: AATTACCTCACTAAAGG
High quality sequence stop: 578.
Location/Qualifiers

FEATURES 1..849
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="C16155 (ML13)"
/sub_species="vulgare"
/db_xref="taxon:112509"

/clone="HV_CEA0012L01f"
 /tissue type="seedling green leaf"
 /lab host="TUC121"
 /clone_lib="Hordeum vulgare seedling green leaf EST
 library HVCDNA0004 (Blumeria challenged)
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
 C.I. 1615 (M1a13) plants were greenhouse grown in the R
 Wise lab at Iowa State University, Ames, IA; 7 day old
 green seedlings were challenged with isolate A27
 (AvrM1a13) of Blumeria graminis f. sp. hordei, and leaves
 were harvested 20 and 24 hr post-inoculation and snap
 frozen; uninoculated leaves were harvested 20 hr
 post-inoculation (Wei, Wise). In the TJ Close lab at the
 University of California, Riverside, total RNA was
 prepared from each sample pool, equal quantities of all
 three RNA pools were combined, poly(A) RNA was purified
 from the mixture, one cDNA library was made, and 1 million
 pla were in vivo excised to give pBluescript SK(-) cDNA
 phagemids (Choi, Close). Phagemids were plated and picked
 at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
 Oates, Rambo, Main). The sequence has been trimmed to
 remove vector sequence and contains a minimum of 100 bases
 of phred value 20 or above. For more details on library
 preparation and sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
 this clone see <http://www.genome.clemson.edu/orders> Also
 see Close TJ, Wing R, Kleinborts A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>).

ORIGIN

Query Match 9.0%; Score 53.8; DB 2; Length 849;
 Best Local Similarity 58.6%; Pred. No. 0.0056;
 Matches 119; Conservative 0; Mismatches 72; Indels 12; Gaps 1;

QY 31 GCCAGACCTGCTGCACGATCTGCCCGTAGAGGTGCTTGAGCCGCGGCC 90
 DB 517 GCCAACAACCTGCTGAGTATCTGTGAGCTTTAAGTTCTTGGAACCTCCCTCC 458
 QY 91 GACGCCAGAACTGCGACGCGGAGCGGCGACCCGCTTGAGACGACGCGGGGTAG 150
 DB 457 GTGGCACAAAACGAGGACGCAAGGGGCGACCGCTGTGATGAAACGCGGGGTAGC 398
 QY 151 CC-----GTCGTGCGACGAGATCCGACCGCTCGAATGAATTGGCGGTGACC 198
 DB 397 CTAGGGGTGCGCTGTGCTGACAGGATCCCTACCGTCTTCACCAACCTGTGTCTCC 338
 QY 199 AGCTCGAAGAGAACTTTGTGCT 221
 DB 337 AGCTTGAGAAATCTTGTGTGT 315

RESULT 12
 CK152637 857 bp mRNA linear EST 05-DEC-2003
 LOCUS FGAS035711 Triticum aestivum FGAS: Talc3 Triticum aestivum cDNA,
 DEFINITION mRNA sequence.
 ACCESSION CK152637
 VERSION CK152637.1 GI:38971884
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (bases 1 to 857)
 Allard, F., Crosby M.L., Danyluk, J., Sudes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,

TITLE Penniket, C., Roach, J.L. and Sathan, F.
 JOURNAL Functional Genomes of Abiotic Stress In wheat and Canola Crops
 COMMENT Unpublished (2003)
 CONTACT: Wm L Crosby
 BIOINFORMATICS
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas_estes@usask.ca

FEATURES
 source
 1..857
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Wheat line CI 14106"
 /db_xref="taxon:4565"
 /lab_host="DH5 alpha"
 /clone_lib="Triticum aestivum FGAS: Talc3"
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
 subtractive hybridization) cDNA library from genotype
 CI14106 cold hardened at 2 C for 21 days and 49 days
 (equal amount of cDNA pooled together before subtraction,
 tester) and subtracted against genotype Norstar cold
 hardened at 2 C for 1 day (24 H (driver). Nitro-pyrole
 anchored oligo-dT priming and non-directional cloning."
 Location/Qualifiers
 1..857

ORIGIN

Query Match 8.9%; Score 53.4; DB 7; Length 857;
 Best Local Similarity 57.6%; Pred. No. 0.0071;
 Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

QY 24 CCGAGCGCCAGACCTGCTGCACGATCTGCCCGTAGAGGTGCTTGAGCCGCTC 83
 DB 452 CTCTATGCGCAAGACTGCTGCACGATCTGCGCCCTTAAGTTCTTGCGAACCTCC 393
 QY 84 GCGGCCAGACGCGCAAACTGCGACGCGGAGCGGCGACCCGCTTGAGACGACGAGC 143
 DB 392 CTTCACAGTGGCAACAGAACGAGCAAGCGGGGCGACCGACTGTGACGAAACAGAGGC 333
 QY 144 GGTGTAGCC-----GTCGTGCGACGAGATCCGACCGCTCGAATGAATTGGCC 191
 DB 332 GGTAGCTTAGAGCGGTGCGCTGTGCAACGGGATCCCGCTCTCCACCAACATGTT 273
 QY 192 GTCGACCACTCGAAGAGAACTTTGTGCT 221
 DB 272 GTCCCTCCAACTTGAGAAATCTTGTGTGT 243

RESULT 13
 BR415712 339 bp mRNA linear EST 24-JUL-2000
 LOCUS MML038.G04000426 ITEC MML Wheat Root Library Triticum aestivum cDNA
 DEFINITION clone MML038.G04, mRNA sequence.
 ACCESSION BR415712
 VERSION BR415712.1 GI:9413558
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (bases 1 to 339)
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,
 Gustafson, P., Herrmann, R.G., Holton, T., Jacquemint, J.M., Jia, J.,
 Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.C., McGuire, P.,

TITLE
Ogilhara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrelle, M., Warburton, M. and Wenzel, G.
International Triticaceae EST Cooperative (ITREC): Production of Expressed Sequence Tags for Species of the Triticaceae

JOURNAL
Unpublished (2000)

COMMENT
Contact: Warburton M
Applied Biotechnology Center, CIMMYT
Apdo. Postal 6-641, 06600 Mexico DF MEXICO
Tel: 52-5-7269091 ext 1381
Fax: 52-5-7267558/59
Email: mwarburton@cgmnet.com
International Triticaceae EST Cooperative (ITREC)
http://wheat.pw.usda.gov/genome.

FEATURES
Location/Qualifiers
1..339
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Atlas"
/db_xref="taxon:4565"
/clone="MM1038.G04"
/issue_type="root"
/dev_stage="8 day old"
/clone_lib="ITREC MML Wheat Root Library"
/note="vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average insert size."

ORIGIN

Query Match 8.8%; Score 52.8; DB 2; Length 339;
Best Local Similarity 58.1%; Pred. No. 0.0096;
Matches 118; Conservative 0; Mismatches 76; Indels 12; Gaps 1;

QY 31 GCCAGACCTGCTGACGATCTGCGCCGCTAGCAGCTGCTTGAAGCCGCTCGCGGCC 90
DB 317 GGCANACCTGCTGACATCTGCGCCCTTAACTGCTCGGCAACCTGCTTCCCA 258
QY 91 GACGCGCAGAACTGCGACGCGACGCGACCCCGCTTGAAGCAGACGCGGCTAG 150
DB 257 GTGGCAGCAAGACGACGCGAGCGGCGGCGAGCTTGAAGCAACGACGCGGTAGC 198
QY 151 CC-----GTGCTGGCAGCGATCCGACCGCTGCGTGAATGCGCTGACC 198
DB 197 CTAGGCGTGCCTGCTGCTCAACGGGAGATCCCACTGCTCCACCACTGTGCTCC 138
QY 199 AGCTGACAGCAACTTGTGCT 221
DB 137 AACTGAGAGATCTGTGTGT 115

RESULT 14
CD922535 528 bp mRNA linear EST 15-JUL-2003
LOCUS G750.103J05F010528 G750 Triticum aestivum cDNA clone G750103J05,
DEFINITION mRNA sequence.
ACCESSION CD922535
VERSION CD922535.1 GI:32770299
KEYWORDS EST
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticaceae; Triticum.
1 (bases 1 to 528)
REFERENCE
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genome programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).

FEATURES
Location/Qualifiers
1..528
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="rectal"
/db_xref="taxon:4565"
/clone="G750103J05"
/issue_type="grain (750 degrees per day after pollination)"
/clone_lib="G750"

ORIGIN

Query Match 8.8%; Score 52.8; DB 6; Length 528;
Best Local Similarity 57.6%; Pred. No. 0.0098;
Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

QY 24 CCGAGCGCCAGCACTGCTGACGATCTGCGCCGCTAGCAGGTGTGCTTGAAGCCGTC 83
DB 270 CTTATAGGCCAATCTGCTGACATCTGCGCCCTTGAAGTGTCTCGCAACCTTCC 211
QY 84 GCGGCGCGCAGCGCGAAGCTGCGACGCGAGGCGCACCCCGCTTGAAGCAGCAGGC 143
DB 210 CTTCCAGTGCACAGAAAGAGACCGACGAGGCGCAGCACTTGCATGAAGCGAGT 151
QY 144 GTGTAGCC-----GTGCTGGCAGCGATCCGACCGCTTGAATGAATGCGC 191
DB 150 TGTAGCTTAGGCGCGCGCCCTGCTGCAAGGCGATCCCACTGCTCCACAGTCTGT 91
QY 192 GTGACCAAGCTGCAACGAACTTGTGCT 221
DB 90 GTCTCCAGCTTGAAGATCTGTGTGT 61

RESULT 15
CK124607/c 771 bp mRNA linear EST 01-MAR-2004
LOCUS CK124607
DEFINITION BS1824109M05 BS1824 Hordeum vulgare subsp. vulgare cDNA clone
MPM2p010M059 5-PRIME, mRNA sequence.
ACCESSION CK124607
VERSION CK124607.1 GI:44807609
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticaceae; Hordeum.
1 (bases 1 to 771)
REFERENCE
AUTHORS Kramer, A., Fellner, T., Possling, A., Radchuk, V., Weschke, W.,
Buerkle, L., and Kersten, B.
TITLE Application of the protein microarray technology for the
identification of expression library derived target proteins for
barley protein kinase CK2
JOURNAL Unpublished (2003)
COMMENT Contact: Birgit Kersten* and Winfriede Weschke**
*Plant Protein Chip Group, Department Lehrich, **Department
Molecular Genetics, Gene Expression Group
*Max-Planck-Institute for Molecular Genetics,
Genetics and Crop Plant Research Gatersleben
*Inneer, 73, D-14195 Berlin, Germany, **Corrensstrasse 3, D-06466
Gatersleben, Germany
Tel: **49(0)30/84131648, **49(0)394825500
Fax: **49(0)30/84131128, **49(0)394825237
Email: **kersten@molgen.mpg.de, **weschke@ipk-gatersleben.de
Insert Length: 771 Std Error: 0.00
Plate: 9 row: M column: 5
Seq primer: PQR65.
Location/Qualifiers
1..771
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Barke"
/db_xref="taxon:4565"
/sub_species="vulgare"
/db_xref="GABI:945637"

```

/db_xref="taxon:112509"
/clone="MPMG2010M059"
/issue_type="embryosac"
/dev_stage="0-10 DAF (days after flowering)"
/lab_host="E. coli, SCS-1/PSK11"
/clone_1lb="BES1824"
/notes="Vector: pQE30NST (AF074376); Site 1: SalI; Site 2:
NotI; 0-10 DAF (days after flowering), cDNA synthesis
using Bluescript II XR cDNA-library construction kit
(Stratagen) with an oligo(dT)-primer containing NotI
restriction site and a SalI adapter (Invitrogen). The main
library of 21500 clones was rearranged into the sublibrary
BES 1824 containing 4100 putative expression clones. Note:
Due to a cloning artefact caused by the kit, in most cases
the SalI site is NOT present, as well as the SalI Adapter
used for cloning. To excise the insert, restriction sites
upstream SalI should be used (e.g. BamHI). Average insert
size is 1 kb. Library generation and sequencing was
granted in context of GABI; data are also accessible at
https://gabi.rzpd.de"

```

ORIGIN

```

Query Match      8.8%; Score 52.8; DB 7; Length 771;
Best Local Similarity 57.6%; Pred. No. 0.01;
Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

QY 24 CCGAGGCGCAGCACTGCTGCAGATCTGCCCGGTAGCAAGTGTGCTTGAAGCCGTC 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 CTCTATGCGCCAAACCTGCTCGACTATCTGTGCGCCCTTAAGGTTCTTACAAACCTCC 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 84 GCGGCGCGAGCGGCAAGTGCAGCGGAGGCGCACCCGCGCTGAGACGAGACGAGCG 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 CTTCCTGTGCGCAAAACGACGACGAGGAGCGACCGACCTGTGATGAAACGACAGGC 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 GGTGTAGCC-----GTGCTGCGCAGCGATCCGACCGTCTCGATGAATTGCC 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 293 GGTGAGGCTTAGGCGTGCCCTGTGCTGACGAGGAGATCCCTAACGCTCCACCAACCTGTT 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 GTGACCAAGTCTGAAACGAGAACTTTGTGCT 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 233 GTCTTCAGCTTAGAAGTATCTTGTGTGT 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: April 7, 2006, 14:04:07
 Job time : 2303.02 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:28:59 ; Search time 98.4702 Seconds
(without alignments)
10776.903 Million cell updates/sec

Title: US-10-611-442-2_COPY_17000_17596
Perfect score: 597
Sequence: 1 aacacgagcgtgtgacgcc.....cacgaacgcgcagcgggcgc 597

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/1 COMB.seq: *
2: /cgn2_6/prodata/1/ina/5 COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/6H COMB.seq: *
6: /cgn2_6/prodata/1/ina/PCITUS COMB.seq: *
7: /cgn2_6/prodata/1/ina/PP COMB.seq: *
8: /cgn2_6/prodata/1/ina/RE COMB.seq: *
9: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.8	13.0	4403765	3	US-09-103-840A-2
2	77.8	13.0	4411529	3	US-09-103-840A-1
3	77.6	13.0	1080	3	US-09-902-540-9688
4	77.6	13.0	13751	3	US-09-902-540-1099
5	75	12.6	792	3	US-09-902-540-3050
6	75	12.6	6713	3	US-09-902-540-793
7	64.8	10.9	456	3	US-09-252-991A-6211
8	64.8	10.9	855	3	US-09-252-991A-6337
9	64.8	10.9	1164	3	US-09-252-991A-6419
10	64.8	10.9	1764	3	US-09-252-991A-6133
11	62.6	10.5	1221	4	US-09-605-703B-2129
12	62.6	10.5	1221	4	US-09-605-703B-2131
13	61.4	10.3	15377	3	US-09-902-540-1116
14	60.8	10.2	966	3	US-09-902-540-2548
15	55.4	9.3	1194	3	US-09-489-039A-4396
16	53.8	9.0	7685	3	US-09-221-017B-1092
17	48.8	8.2	1086	3	US-09-107-433-2203
18	48.8	8.2	1095	3	US-09-107-433-2203
19	48.8	8.2	5558	3	US-08-961-527-103
20	47.4	7.9	400	3	US-08-956-171E-1342
21	47.4	7.9	400	3	US-08-781-986A-1342
22	46.8	7.8	3758	3	US-09-902-540-6742
23	46.8	7.8	3760	3	US-09-902-540-543
24	45.2	7.6	738	3	US-09-252-991A-13548

C 25	45.2	7.6	1053	3	US-09-252-991A-13694	Sequence 13694, A
C 26	45.2	7.6	1164	3	US-09-252-991A-13896	Sequence 13896, A
C 27	45.2	7.6	1293	3	US-09-252-991A-13447	Sequence 13447, A
C 28	44.8	7.5	1990	3	US-09-252-991A-3181	Sequence 3181, Ap
C 29	44.8	7.5	1548	3	US-09-252-991A-2688	Sequence 2688, Ap
C 30	44.8	7.5	2523	3	US-09-252-991A-2990	Sequence 2990, Ap
C 31	43.2	7.2	822	3	US-09-252-991A-5647	Sequence 5647, Ap
C 32	43.2	7.2	1428	3	US-09-252-991A-5676	Sequence 5676, Ap
C 33	43.2	7.2	1998	3	US-09-252-991A-5524	Sequence 5524, Ap
C 34	43.2	7.2	2187	3	US-09-252-991A-5616	Sequence 5616, Ap
C 35	42.8	7.2	8931	3	US-09-028-934-28	Sequence 28, Ap1
C 36	41.8	7.0	1162	3	US-09-395-674B-3	Sequence 3, Ap1
C 37	41.2	6.9	1800	3	US-09-252-991A-15332	Sequence 15332, A
C 38	41.2	6.9	1800	3	US-09-252-991A-15332	Sequence 2716, Ap
C 39	41.2	6.9	2133	3	US-09-252-991A-2716	Sequence 15284, A
C 40	41.2	6.9	2319	3	US-09-252-991A-15284	Sequence 15396, A
C 41	41.2	6.9	2427	3	US-08-804-227C-7	Sequence 7, Ap1
C 42	41.2	6.9	44377	2	US-08-804-198-1	Sequence 1, Ap1
C 43	41.2	6.9	792	3	US-09-252-991A-9729	Sequence 9729, Ap
C 44	41	6.9	792	3	US-09-252-991A-9687	Sequence 9687, Ap
C 45	41	6.9	1512	3	US-09-252-991A-9687	Sequence 9687, Ap

ALIGNMENTS

RESULT 1
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 13.0%; Score 77.8; DB 3; Length 4403765;
Best Local Similarity 61.7%; Pred. No. 2e-09;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY	26	CGAGCCCGACACTCTCTGACAGATTCGCCCGGTGACAGGTCTGCTTGAAGCCCTGCG	85
DB	3184316	CGCCGCGACGACCTGCTCGAGAGATCTCCGCGTCATGATGTCGACGATCAATCCGCCCT	3184375
QY	86	GGCCCGACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	145
DB	3184376	GGCCAGTGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	3184435
QY	146	TGTAGCCCTGCTGCGACGACGATCCGACGCTCTGATGTAATGACCGTCGACGACGTCGA	205
DB	3184436	TGTGCGCGCGCGGATGACGATCAGACGACGATTCGAACATGATGACGATGACGCGCGC	3184495
QY	206	ACAGGACCTTTGCTGCTGCTGCGC	226
DB	3184496	ACAACTCTTTCGAGCTGCGC	3184516

RESULT 2

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 13.0%; Score 77.8; DB 3; Length 4411529;
Best Local Similarity 61.7%; Pred. No. 2e-09;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 26 CGAGCCGACGACTGCTGCGACGATCTGCGCCGGTAGAGGTGTCCTTGAAGCCGCTGCGC 85
DB 3190207 CGCCGCGACGCGACTGCTGCGAGATCTCCGCGTAGAGGTGTCCTTGAAGCCGCTGCGC 3190266
QY 86 GCGCCGACGCGGAGACTGCGACGCGGAGGCGACCCCGCTGAGACGAGCGGCGG 145
DB 3190267 GCGCAGTGGCACAGACGCGGAGGAGCCATGCGCAGCCGCTTGGCAGGAAATGCAAGCG 3190326
QY 146 TGTAGCCGCTGCGGAGATCCGACCGCTCTCGATGAATGCGCGTGAACCACTGCA 205
DB 3190327 TGTTCGCGCGGAGTAGGCGATGACGACGATTTGGAACATGTCACCGTGAACGCGCCG 3190386
QY 206 ACAGGAACCTTTGCTGCTGCGC 226
DB 3190387 ACAACGCTTTGAGTCTGCGC 3190407

RESULT 3

US-09-902-540-9688/c
Sequence 9688, Application US/09902540
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 9688
LENGTH: 1080
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-9688

Query Match 13.0%; Score 77.6; DB 3; Length 1080;
Best Local Similarity 61.3%; Pred. No. 4.2e-10;
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 34 AGCACTGCTGACGATCTGCGCCGATAGCAGGTGCTTGAAGCGCGTGGCGCCGAC 93
DB 431 AGTACTGCGCCACGATTTGCGCGGCGGTCAAGTGGCGGTGAAGCCCATCTGCGCGGTG 372

QY 94 GCGGAGAACTGGACGCGGAGGCGACCCCGCTGAGACGACGAGCGGTGACCG 153
DB 371 GCGGAGAAAGCGAGCCGACATGCGCGACCGCGCTGCTGTGATGACACCGTGGCGCGG 312
QY 154 TCGTGGGACGAGATCCGACCGCTCTCGATGAATGCGCGTGAACCGAGTGAACAGAAC 213
DB 311 CCTTGAACCGCATACGACGAGTTTCATGATGTCGCCGTCGTCAGTCCAGCAGCAGC 252
QY 214 TTTGTGCTGCTGCGCTTCCCTGCGT 237
DB 251 TTGTGCGTGAAGCGGTGCTGCTG 228

RESULT 4

US-09-902-540-1099/c
Sequence 1099, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1099
LENGTH: 13751
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-1099

Query Match 13.0%; Score 77.6; DB 3; Length 13751;
Best Local Similarity 61.3%; Pred. No. 7e-10;
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 34 AGCACTGCTGACGATCTGCGCCGATAGCAGGTGCTTGAAGCGCGTGGCGCCGAC 93
DB 5590 AGTACTGCGCCACGATTTGCGCGGCGGTGAGTGGCGGTGAGACCCCATCTGCGCGGTG 5531
QY 94 GCGGAGAACTGGACGCGGAGGCGACCCCGCTGAGACGACGAGCGGTGACCG 153
DB 5530 GCGGAGAAAGCGAGCCGACATGCGCGACCGCGCTGCTGTGATGACACCGTGGCGCGG 5471
QY 154 TCGTGGGACGAGATCCGACCGCTCTCGATGAATGCGCGTGAACCGAGTGAACAGAAC 213
DB 5470 CCTTGAACCGCATACGACGAGTTTCATGATGTCGCCGTCGTCAGTCCAGCAGCAGC 5411
QY 214 TTTGTGCTGCTGCGCTTCCCTGCGT 237
DB 5410 TTGTGCGTGAAGCGGTGCTGCTG 5387

RESULT 5

US-09-902-540-3050/c
Sequence 3050, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825


```
; SEQ ID NO 3050
;
; LENGTH: 792
;
; TYPE: DNA
;
; ORGANISM: Myxococcus xanthus
;
US-09-902-540-3050
```

Query Match	12.6%;	Score 75;	DB 3;	Length 792;
Best Local Similarity	58.9%;	Pred. No. 1.8e-09;		
Matches 129;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0;

Oy 3 AACGAGCCCTGTGAGCGCCGGGCCCCGAGCGCCAGCACTCTGTCGACGATCTTCGCCGGTACG 62
 Db 459 CTCGTTCTTGGCAACCTCGCGGTTACCCCGTGGACCTTGGGCCACAACTTCCTGGGGGT 400
 Oy 63 CAGGTGTGCTTGTAGGCGCGCTCGCGGCCGACGCGCGAAGACTGGACGCGAGAGCGCACCC 122
 Db 399 CAGGTGTGCGCTTGTAGGCCCATGTGTGTCGCTCATGTGAGAGCGCACGGCCATGCGCGAGCC 340
 Oy 123 CGCCTGAAACAGAGCGCAGGCGCGGTGTAGCGCCGTGTGTCGTCGAGCGGATCCGACCGCTTCGAT 182
 Db 339 CACCTGCGATGACAGCGACACMAAGTCTCTGGCGTCTCTCGTGGGATGTAGACGGATTCGAT 280
 Oy 183 GAAATTGCGGTGACCGAGCTCGAAGCAGAACTTTGCGT 221
 Db 279 GTAGCGCCCGTCCCGGTCTTCAACCGGTACTTGATGCT 241

```

RESULT 6
US-09-902-540-793
; Sequence 793, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(1584)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 793
; LENGTH: 6713
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-793

```

	Query Match	12.6%	Score 75	DB 3	Length 6713
	Best Local Similarity	58.9%	Pred. No. 2.7e-09		
	Matches	129	Conservative	0	Mismatches 90
			Indels	0	Gaps
QY	3	CACGAGCCTGTTGACGCGGCGCCGAGCGCGCAGACCTTGTCGACGATCTGCGCCGGTAG	62		
Db	336	CTCGTTCCTTGGAACCTGCGCGTTCACGCGGTGACCTGGGACCAACATCTCGTGGGGGT	395		
QY	63	CAGGTGTGCTTGAAGCGCGTGGCGGCCGCGACGCGGAGAACTGTGACGCGAAGCGGACCC	122		
Db	396	CAGGTGCGCTTGAGCCCACTGATGTCCCGTCAATGACGAAGCCGCAAGCCATGCGAGGC	455		
QY	123	CGCCTGACGACGACGCAAGCGGTGTAGCCGTCTGTGCGGACCGGATCCGACCGTCTCAT	182		
Db	456	CACCTGCGTGAACGACGACAAAGTCTCTGGCGGTCTCTGGGGGCAATGTAGACGGAATTCGAT	515		
QY	183	GAAATTCGCGCTGCACCACTCGAAGAAGAACTTTTCGT	221		
Db	516	GTAGCGCCGCTCCGCGTCTTCACCGGACTTGATGGT	554		

RESULT 7
US-09-252-991A-6211
; Sequence 6211, Application US/09252991A

Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERGININOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252.991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074.788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094.190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 6211
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-6211

Query Match	10.9%	Score 64.8	DB 3	Length 456
Best Local Similarity	60.0%	Pred. No. 5.6e-07		
Matches 108; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0

Qy	36	CACTCTCTCAAGATCTGGCCCGGAGGAGGTGTCGTGAGGCGCGTGGCGGCCAAGC	95
Db	144	CACTTGGCCGATCACCCTGGCGCGCGGTGAGGTCCGTGTTGAACCTTGCCTTGGCGGTGA	203
Qy	96	GCAGAACTGGACGCGAAGGCGACCCCGCTGAGACGAGCGAGGCGGTGAGCCGTC	155
Db	204	GCAGAACTGGCAATCGAGGCGCGACCCCGCTTGGAGGACACGACAGGGGTGCCGCTCC	263
Qy	156	GTGGCGACGGAATCCGACACCGTCTCGATTAAATTCCCGTCGACCAAGTCTGAAACAGAACTT	215
Db	264	GCCCTTGGGGAATGTACACGGTCTCGACGCACTGCGGAGCCACCCGGAACCAACCACTT	323

```

RESULT 8
US-09-252-991A-6337/C
; Sequence 6337, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6337
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-6337

```

	Query Match Best Local Similarity	10.9% 60.0%	Score 64.8 Pred. No. 6.4e-07	DB 3.1	Length 855
	Matches 108; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0
Qy	36	CACCTGCTGACGATCTCGCCCGGTACAGGTGTGCTTGAAGGCGTGCAGGCCCGACGCG	95		
Db	436	CACCTGGCCGATACCTCGGGGGGTGAGGTTCCTGTGAACCTCTTCCCGGTGGA	377		
Qy	96	GCAGAACCTGACGCGAAGGCGACCCCGCCTTGAGACGACGACGAGCGGTGTAGCCGTC	155		
Db	376	GCAGAACCTCAATCCAGGGCGACCCCGCCTTGAGAGACGACGACGAGGTGTCGCGCTCC	317		
Qy	156	GTGGCCACCGATCCGCGACCGTCTTCGATGAAATTGCCCTCCACACGATCGAACAAGAACTT	215		
Db	316	GCCCTGGGGATGTACACGGTCTTCGACGCAATGCCCGACCGCACCCGGAACACCCACTT	257		

RESULT 9

US-09-252-991A-6419/c
; Sequence 6419, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6419
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6419

Query Match 10.9%; Score 64.8; DB 3; Length 1164;
Best Local Similarity 60.0%; Pred. No. 6.8e-07;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 36 CACCTGCTGCACGATCTCGCCGGTAGACAGGTGTGCTTGAGCCGCTCGCCGCCGACGC 95
DB 456 CACCTGGCCGATCACTCTGGCGCGGTGAGGTGCTGTTGAACCTGCTTCCGGTGA 397
QY 96 GCAGAACTGGACCGCAGAGCGCAACCCGCTGAGACAGACGAGCGGTGACCGTC 155
DB 396 GCAGAACTGGACCGCAGAGCGCAACCCGCTGAGACAGACGAGCGGTGACCGTC 337
QY 156 GTGGCGACGATCCGACCGCTGCTGATGAATTTGCGTGCACGCTGAAACGAACTT 215
DB 336 GCCCTGGGGATGTACACGCTCTGACGCACTGCCGACCGCACCCGACCCACTT 277

RESULT 10

US-09-252-991A-6133
; Sequence 6133, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6133
; LENGTH: 1764
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6133

Query Match 10.9%; Score 64.8; DB 3; Length 1764;
Best Local Similarity 60.0%; Pred. No. 7.3e-07;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 36 CACCTGCTGCACGATCTCGCCGGTAGACAGGTGTGCTTGAGCCGCTCGCCGCCGACGC 95
DB 1525 CACCTGGCCGATCACTCTGGCGCGGTGAGGTGCTTGAACCTGCTTCCGGTGA 1584
QY 96 GCAGAACTGGACCGCAGAGCGCAACCCGCTGAGACAGACGAGCGGTGACCGTC 155

DB 1585 GCAGAACTGCATTCACGCGCGACGCCCTGGGAGACACGACAGGGTCCGCTCC 1644
QY 156 GTGGCGACGATCCGACCGCTCTGATGAATTTGCGTCCAGCAGCTCGAAGAACTT 215
DB 1645 GCCCTGGGGATGTACACGCTCTGACGCACTGCGGAGCGCACCGGACCACTT 1704

RESULT 11

US-09-605-703B-2129/c
; Sequence 2129, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zeider, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; PRIOR FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 2129
; LENGTH: 1221
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1198)
; OTHER INFORMATION: RXA01186
US-09-605-703B-2129

Query Match 10.5%; Score 62.6; DB 4; Length 1221;
Best Local Similarity 59.1%; Pred. No. 2.4e-06;
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACCTGCTGCACGATCTCGCCGGTAGACAGGTGTGCTTGAAGCCGCTCGCGGCCGACG 94
DB 566 GAACCTATCCACATCTCACCCGATGAAGTTGATGAGTCCAAACCCCTGACCAATTG 507
QY 95 CGCAGAACTGGACCGCAGAGCGCAACCCGCTGAGACGAGACGAGCGGTGACCGT 154
DB 506 CACGAAATGGGACCGCATGCTCCGACCAAGCTTGCAGAAATTCACAGCTGAGCGAT 447
QY 155 CGTGGCAGCGATCCGACCGCTCTGATGAATTTGCCGTGACCAAGCTCGAACAGAACT 214
DB 446 CTGAATAGCGCATGAGAACAGACTCAAGTGAAGTCCATCATGAGCTTCATTAAGCTT 387
QY 215 T 215
DB 386 T 386

RESULT 12

US-09-605-703B-2131/c
; Sequence 2131, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zeider, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; PRIOR FILING DATE: 2000-06-27

PRIOR APPLICATION NUMBER: 60/142,764
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: 60/152,318
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 2934
SEQ ID NO 2131
LENGTH: 1221
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(1198)
OTHER INFORMATION: RXA01186
US-09-605-703B-2131

Query Match 10.5%; Score 62.6; DB 4; Length 1221;
Best Local Similarity 59.1%; Pred. No. 2.4e-06;
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACCTGTCACGATCTGCGCCGCGTACGAGTGTGCTTGAAGCGCGCGCCGACG 94
DB 566 GAACCTGATCCAGATCTCACCATTGAAGGTTAGGTCAAAACGCGCTTGACAGTTG 507
QY 95 CGCAGAACTGGACCGGAGGCGCACCCGCTGAGACGACGCGCGGTAGCCGT 154
DB 506 CACAGATGGGACGCGCATCCGCAACAGCTCGGAGAAATACACAGGTGAGCGAT 447
QY 155 CGGCGGAGCGATCCGCGCTTCGATGAATTCGCGTGCACGCTGCAACGACT 214
DB 446 CTGAATAGCGCATGAGAACTCAAGCAAGTGCATCATGAGCTTCATTAAGCTC 387
QY 215 T 215
DB 386 T 386

RESULT 13
US-09-902-540-1116/C
Sequence 1116, Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1116
LENGTH: 15377
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-1116

Query Match 10.3%; Score 61.4; DB 3; Length 15377;
Best Local Similarity 48.6%; Pred. No. 8e-06;
Matches 198; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 3 CACGAGCTGTGACGCGGCGCCGACGCGCGACCTGTGACGATTCGCGCGGTAG 62
DB 11342 CACCGGCGGTCGCTTCTCCGACCTGAGACCTGTCCATATCTCCAGGCTG 11283
QY 63 CAGGTGCGCTTGAAGCGCGCGCGACGCGCGAGAACTGGCAAGCGGAGCGGACCC 122
DB 11282 GAGGTTCCTTTGAAGCCCACTTCCCGTCATGCAAGAGTGCACGCGGCAAGCC 11223
QY 123 CGCTGAGACGAGACGCGCGGTGA---GCCGTGTGGCGACGAGTCCGACCGTCTC 179
DB 11222 CACCTGGCTGAGAGCGCAATGACGTACTTTCATGAAAGATGGGATCGGACGCTTC 11163

QY 180 GATGAATTCGCGTCGACGATCTGAAACGAACTTGTCTGTCTGCTTCCCTGTGCG 239
DB 11162 GATAGCTCCGCGCAGCGGCGAGTGAAGAGTACTTCAAGAACCGCTGTCGCGCGCG 11103
QY 240 ACTGCGCTCGGCGAGGATGCAACGAGAGTGTGTCAGTTGCCCTAGTCTTACGCGTGT 299
DB 11102 GCGCTGACGATGTTCCAGCTTCGCGCATCTCCGCTGGGACGAGAGTGTCTCCACGCG 11043
QY 300 GCGCGATCTTTCGCTGCGGATTAAGCTTGTGAAAGATGTGCGTGTCTTCCGTTGCGC 359
DB 11042 GCGCGGACCTGACGCGGCGGAGCGAGCTCTTCCAGACTGCGCACCGTGGCGAAGAC 10983
QY 360 GCGGAGCGCTCCGCGACGCTCGGAGACGACGAGTGAAGACCGGAGG 406
DB 10982 GCGCGGAAACCTTGTGAGACGCGCGGCGGCGAGCGGAGCGCG 10936

RESULT 14
US-09-902-540-2548/C
Sequence 2548, Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 2548
LENGTH: 966
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-2548

Query Match 10.2%; Score 60.8; DB 3; Length 966;
Best Local Similarity 52.3%; Pred. No. 6.5e-06;
Matches 134; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 3 CACGAGCTGTGACGCGGCGCCGACGCGCGACCTGTGACGATTCGCGCGGTAG 62
DB 342 CACCGGCGGTCGCTTCTCCGACCTGAGACCTGTCCATATCTCCAGGCTG 283
QY 63 CAGGTGCGCTTGAAGCGCGCGCGCGCGGACGCGGAACTGGCAAGCGGAGCGGACCC 122
DB 282 GAGGTTCCTTTGAAGCCCACTTCCCGTCATGCAAGAGTGCACGCGGCGCACGCGC 223
QY 123 CGCTGAGACGAGACGCGCGCGGTGAGCGCGTGTGCGGAGATCCGACCGTTCGAT 182
DB 222 CACCTGCGTGAAGAGCGGAGTACGTACTTCTCATGAAAGTGGAGTCCGACGCTTC 163
QY 183 GAAATTCGCGTGAACGAGTCAACGAACTTGTGCTGTGCTTCCCTGTGAGACT 242
DB 162 GATAGCTCGCGCGGCGGAGTCAAGAGGTAATTCAAGAGCGGTGTGCGCGGCGG 103
QY 243 GCGCTCGCGGAGGTC 258
DB 102 GCGCTGACGATTCG 87

RESULT 15
US-09-489-039A-4396/C
Sequence 4396, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Bretton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 4396
LENGTH: 1194
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4396

Query Match

9.3%; Score 55.4; DB 3; Length 1194;

Best Local Similarity 59.0%; Pred. No. 0.00015;
Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY	37	ACCTGCTGCACGATCTCGCCCGTAGCAGGTGTGCTTGAGGCCGTGCGGCCCGACGCG	96
DB	494	ACCTGCCGATTAATTTCGATACCCGACGTTGCGTTAAAGCCCTGCTGCGCGGTGGAG	435
QY	97	CAGAACTGGCAGCGGAGGCGCACCCCGCTGAGACGAGACGCAAGCGGTGTAGCCCTCG	156
DB	434	CAGAACTTCACCTCCAGCGCGCAGCTTACTTGGAGAGAGACGCAAGGTTGCGCGGTCT	375
QY	157	TGGGAGGAGATCCGACCGCTCTCGATGAATTGACCGTCGAC	197
DB	374	TCTTCCGGATTAAGACGCTCTCGACGCGCTGATGCCAAC	334

Search completed: April 7, 2006, 02:59:01
Job time : 102.47 secs


```
Db 17180 ATGAATTTGCGTGCAGCTCGAACAAGAACTTTGCTGCTTCCCTGTGCGA 17239
Qy 241 CTGCGCTCGCGAGAGGTTCAGACGAAGTCTTCAGATTGCCCTAGTGTCTTCAAGGTGTGG 300
Db 17240 CTGCGCTCGCGAGAGGTTCAGACGAAGTCTTCAGATTGCCCTAGTGTCTTCAAGGTGTGG 17299
Qy 301 GCCGAGCTTTGCGCTCGCGAATAAAGCTTGTGAGAAATGTCGCTGCTGCGCTTCCGCG 360
Db 17300 GCCGAGCTTTGCGCTCGCGAATAAAGCTTGTGAGAAATGTCGCTGCTGCGCTTCCGCG 17359
Qy 361 CCGACGCGCTCGCGAGCTCGAGACGACAGTTCGAGACCGACCGCGCGAGGTGTCT 420
Db 17360 CCGACGCGCTCGCGAGCTCGAGACGACAGTTCGAGACCGACCGCGCGAGGTGTCT 17419
Qy 421 CGTCCGCGAATGGGTAGACCAAGACCTGCGGCGAGAGTTCAGATGTCACCACTTATCA 480
Db 17420 CGTCCGCGAATGGGTAGACCAAGACCTGCGGCGAGAGTTCAGATGTCACCACTTATCA 17479
Qy 481 CGGTGCAAGACGTCAATTGTCAAGTGCACAGAGAGAGCTGACGATGATGCT 540
Db 17480 CGGTGCAAGACGTCAATTGTCAAGTGCACAGAGAGAGCTGACGATGATGCT 17539
Qy 541 CTGCTGTCTTCCGCAATATAGCGTTCGAGCTGCAATTCACGACGCGCGCGC 597
Db 17540 CTGCTGTCTTCCGCAATATAGCGTTCGAGCTGCAATTCACGACGCGCGCGC 17596
```

RESULT 2
US-10-156-761-2608/c
Sequence 2608, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, JUN
APPLICANT: ISHIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 2608
LENGTH: 1104
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1104)
US-10-156-761-2608

Query Match 12.9%; Score 77.2; DB 6; Length 1104;
Best Local Similarity 61.4%; Pred. No. 9, 5e-13;
Matches 124; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy 14 TGACGCCGGGCGCCGAGCCGACACCTGCTGACGATCTCGCCGGTAGCAGGTTCCT 73
Db 430 TGAGGCGCCGCGATGCGCTGCAAGATCTGTGACATCTCGCCGGTTCGACAGGTTCCT 431
Qy 74 TGAGGCGCCGCGCGCCGAGCGGAGAACTGGACGCGAAGGCGACCCCGCTGAGACG 133
Db 430 CCAAGCGCGCTGCTGCGGTGGGCGAAGAGCGGCAAGTTATTCACGACCCCGCTGGAGC 371
Qy 134 AGACGCAAGCGGTGTAGCGGTGTGCGACGATCGACCGCTTCTGATGAATTCGCGT 193
Db 370 TGATGCATATGTCACCGCGTCCGGTAGCGCATGAGACGCACTGAGAGAGCGTCCGT 311
```

```
Qy 194 CGACCACTTCGAACAGAACTT 215
Db 310 CGAACAGCGCCGACAGCGTCTT 289
```

RESULT 3
US-10-156-761-1

Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, JUN
APPLICANT: ISHIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 12.9%; Score 77.2; DB 6; Length 9025608;
Best Local Similarity 61.4%; Pred. No. 2, 2e-12;
Matches 124; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy 14 TGACGCCGGGCGCCGAGCCGACACCTGCTGACGATCTCGCCGGTAGCAGGTTCCT 73
Db 3221186 TGACGCCGGGCGATCCGCTCCAGATCTGTGACGATCTCCCGCTGACAGGTTCCT 3221245
Qy 74 TGAGGCGCCGCGCGCCGAGCGGAGAACTGGACGCGAAGGCGACCCCGCTGAGACG 133
Db 3221246 CCAAGCGCGCTGCTGCGGTGGGCGAAGAGCGGCAAGTTATTCACGACCCCGCTGGAGC 3221305
Qy 134 AGACGCAAGCGGTGTAGCGGTGTGCGACGATCCGACCGCTTCTGATGAATTCGCGT 193
Db 3221306 TGATGCATATGTCACCGCGTCCGGTAGCGCATGAGACGCACTGAGAGAGCGTCCGT 3221365
Qy 194 CGACCACTTCGAACAGAACTT 215
Db 3221366 CGAACAGCGCCGACAGCGTCTT 3221387
```

RESULT 4
US-10-470-565-1/c

Sequence 1, Application US/10470565
Publication No. US20040126870A1

GENERAL INFORMATION:
APPLICANT: Societe des Produits Nestle S.A.
TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
FILE REFERENCE: 80290/WO
CURRENT APPLICATION NUMBER: US/10/470,565
CURRENT FILING DATE: 2003-07-29
PRIOR APPLICATION NUMBER: EP 01102050.0
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1

LENGTH: 2256646
TYPE: DNA
ORGANISM: Bifidobacterium longum
US-10-470-565-1

Query Match 12.4%; Score 73.8; DB 7; Length 2256646;
Best Local Similarity 60.0%; Pred. No. 2e-11;
Matches 123; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 26 CGAGCGGACGACCTGTCGACGATCTCGCCCGGTAGCAGTGTGCTTGAAGCCGCTCCG 85
DB 1885220 CGGCCACACGCGACCTGTCGATTTATTTACCGGTGACATGTTACGGGTACGCGCAAGCT 1885161
QY 86 GGGCCGACGCGGAACTGCGACGCGGAGCGGACCCGCTGAGACGACGAGCGG 145
DB 1885160 TGGCGGTGGCGGACGAGGCGGACGCTCCATGCCGACGCGACCTGCGAGGAAATGACAGCG 1885101
QY 146 TGTAGCCGTCGTGGCGGACGATCCGACCGCTCTGATGAAATTGCCCTGACCAAGCTCGA 205
DB 1885100 TGGTGGGGTGGGGGTATCGCATAGACCGGATTCGATGAGGAGCGCGTCAACAGCTTCC 1885041
QY 206 ACAGAACTTGTCTGCTGCTTCC 230
DB 1885040 ACAGGCTTGTGATGATGCTGCTTCC 1885016

RESULT 5

US-09-738-626-2214/C
Sequence 2214, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: AMDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2214
LENGTH: 1098
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2214

Query Match 10.5%; Score 62.6; DB 3; Length 1098;
Best Local Similarity 59.1%; Pred. No. 2.1e-08;
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACTGCTGCACGATCTCCGCCGGTAGACAGTGTGCTTGAAGCCGCTGCGCGCGACG 94
DB 466 GAACCTGATCCAGATCTCACCGATTGAAGGTTACGGTCCAAACCGCCCTGACCAAGTTG 407
QY 95 CGAGAACTGGACGCGGAGGCGGACCCGCTGAGACGACGAGCGGCTTACCGCT 154
DB 406 CACAGATGGGACGCGGACGCGGACGCGCTGCGAAGAAATACACAGGTTGAGCGAT 347
QY 155 CGTGGGACGATCCGACCGCTCTCGATGAAATTGCCCTGACCACTCGAAGGAACT 214

DB 346 CTGAATAGGCGATGAGAACAGACTCAAGAAAGTGCATCATGAGCTTCATACGCTCT 287
QY 215 T 215
DB 286 T 286

RESULT 6

US-09-738-626-1
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: AMDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 10.5%; Score 62.6; DB 3; Length 3309400;
Best Local Similarity 59.1%; Pred. No. 4.4e-08;
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACTGCTGCACGATCTCCGCCGGTAGACAGTGTGCTTGAAGCCGCTGCGCGCGACG 94
DB 2135086 GAACCTGATCCAGATCTCACCGATTGAAGGTTACGGTCCAAACCGCCCTGACCAAGTTG 2135145
QY 95 CGAGAACTGGACGCGGAGGCGGACCCGCTGAGACGACGAGCGGCTTACCGCT 154
DB 2135146 CACAGATGGGACGCGGACGCGGACGCGCTGCGAAGAAATACACAGGTTGAGCGAT 2135205
QY 155 CGTGGGACGATCCGACCGCTCTCGATGAAATTGCCCTGACCACTCGAAGGAACT 214
DB 2135206 CTGAATAGGCGATGAGAAAGACTCAAGAAAGTGCATCATGAGCTTCATACGCTCT 2135265
QY 215 T 215
DB 2135266 T 2135266

RESULT 7

US-09-974-300-6756/C
Sequence 6756, Application US/0974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 6756
LENGTH: 672
TYPE: DNA
ORGANISM: Bacillus clausii
US-09-974-300-6756

Query Match 9.4%; Score 59.4; DB 3; Length 672;
Best Local Similarity 57.1%; Pred. No. 1.8e-07;
Matches 108; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 33 CAGACCTGCTGCACGATCTGCCCCGGTAGAGGTGCTTGAGCCGCTGCGGCCCGCA 92
DB 446 CATATCTGCTGCACAAATTTGACCGCTGGAACAAATCGCGGTTTTTAAGCAACAGCACT 387
QY 93 CGGCGAGAACTGGACCGGGAAGGCGACCCCGCTGAGACGACGAGCGGTGTAGCC 152
DB 386 CGCACAAAGCTACGACCAATGTTGACGCCCACTGTGTGTAAAGCAACCGAGAGCC 327
QY 153 GTGCTGGCAGCATCCGACCGTCTCGATGAAATTCGCTGACCACTCGAAACAGAA 212
DB 326 GTACTTATGGCGCATTAAGCAACGCTCTCAATTAATTCGCTTGAGACGAGAAAGAAA 267
QY 213 CTTTGTCT 221
DB 266 TTATATCT 258

RESULT 8

US-10-450-763-18200/C
Sequence 18200, Application US/10450763
Publication No. US20050196754A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 790CIP3/US
CURRENT APPLICATION NUMBER: US/10/450,763
CURRENT FILING DATE: 2003-06-11
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO: 18200
LENGTH: 2493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIMILAR
LOCATION: (403)..(1233)
OTHER INFORMATION: 95% homologous to Escherichia coli similar to accession
US-10-450-763-18200

Query Match 9.4%; Score 56.4; DB 9; Length 2493;
Best Local Similarity 59.3%; Pred. No. 1.6e-06;
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 36 CACCTGTCAGCATCTGCGCCGGTAGACGATGCTTGAGCCGCTGCGGCCCGACGC 95
DB 660 CACCTGCGCGATATTTCCGACACCCGAGGTGGTTAAAGCCCTGTGGCGGTGA 601
QY 96 GCAGAACTGGACCGGGAAGGCGACCCCGCTTGAAGCAGACGACGCGGTGTAGCCGTC 155
DB 600 ACAGAAATTACATCCAGCGCACCCCACTGCGAAGAGACGACGAGCGGTGCAACGCTC 541

QY 156 GTGGCAGCGATCCGACCGCTCTGATGAATTCGCTGAC 197
DB 540 GTCTCGGATATACCGCTTTCAGACGCGCTGATGCCAAC 499

RESULT 9

US-10-450-763-25444/C
Sequence 25444, Application US/10450763
Publication No. US20050196754A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 790CIP3/US
CURRENT APPLICATION NUMBER: US/10/450,763
CURRENT FILING DATE: 2003-06-11
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO: 25444
LENGTH: 2493
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIMILAR
LOCATION: (403)..(1233)
OTHER INFORMATION: 95% homologous to Escherichia coli similar to accession
US-10-450-763-25444

Query Match 9.4%; Score 56.4; DB 9; Length 2493;
Best Local Similarity 59.3%; Pred. No. 1.6e-06;
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 36 CACCTGTCAGCATCTGCGCCGGTAGACGATGCTTGAGCCGCTGCGGCCCGACGC 95
DB 660 CACCTGCGCGATATTTCCGACACCCGAGGTGGTTAAAGCCCTGTGGCGGTGA 601
QY 96 GCAGAACTGGACCGGGAAGGCGACCCCGCTTGAAGCAGACGACGCGGTGTAGCCGTC 155
DB 600 ACAGAAATTACATCCAGCGCACCCCACTGCGAAGAGACGACGAGCGGTGCAACGCTC 541
QY 156 GTGGCAGCGATCCGACCGCTCTGATGAATTCGCTGAC 197
DB 540 GTCTCGGATATACCGCTTTCAGACGCGCTGATGCCAAC 499

RESULT 10

US-09-974-300-2461/C
Sequence 2461, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 2461
LENGTH: 594
TYPE: DNA

ORGANISM: Bacillus licheniformis
US-09-974-300-2461

Query Match 9.3%; Score 55.6; DB 3; Length 594;
Best Local Similarity 58.4%; Pred. No. 2.4e-06;
Matches 97; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 56 CCGGTACAGAGTGTGGTGGAGCGCCGCGAGAGTGGAGCGAGG 115
DB 426 CCGGTTCATGTTCCGTTTCAGCCGCGAGGCTCATTAAGTACGCGATCC 367
QY 116 CGGACCCCGCTGAGAGAGAGCGAGCGGTGTAGCCGTGCGAGAGATCCGACCG 175
DB 366 GGCCTCCACCTGTGTGAGAACAGAAATGCGATATGTGCGCGATTAATACGG 307
QY 176 TCTGATGAATTTCCCTGACACGCTCGAAGAGAACTTTGTGT 221
DB 306 TCTGATGTGTAGCCGTGATGAGATGAATTAATACTTATGT 261

RESULT 11

US-10-194-163-1092
Sequence 1092, Application US/10194163
Publication No. US20020172976A1

GENERAL INFORMATION:

APPLICANT: Rose, Bruce Carter
TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: PALO ALTO

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/194,163

FILING DATE: 04-Nov-2002

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Basu, Shantanu

REGISTRATION NUMBER: 43,318

REFERENCE/DOCKET NUMBER: 529282000101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5995

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 1092

SEQUENCE CHARACTERISTICS:

LENGTH: 7685 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..7685

SEQUENCE DESCRIPTION: SEQ ID NO: 1092

US-10-194-163-1092

Query Match 9.0%; Score 53.8; DB 5; Length 7685;
Best Local Similarity 55.7%; Pred. No. 1e-05;
Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 37 ACCTGCTGACGATCTGCCCCGTAGAGGTGTGCTTGAAGCCGTGCGGCCGAGCCG 96
DB 743 ATCTGTTGAGATCTGCGTGTGAGAGATTCCTGTCATCTCTGTTTTCGGTATG 802
QY 97 CAGAACTGACAGCCGAGAGGCGCACCCCGCTGAGAGAGAGAGGCGGTGTAGCCGTG 156
DB 803 CAGAAAGGAGATGCTCATTTTTCGACACGACCTGAGAGAGATACAGAGCTTAGCCCTTGC 862
QY 157 TGGCAGAGATTCGACACCGCTGTGATGAATTCGCTGACAGACTTGAA CAGAACTTT 216
DB 863 CCTCAGAGATAGTACAGATTCACGAAAGCCCTTCTCTACGGGGAAGATACCTTC 922
QY 217 GTGT 221
DB 923 TTGT 927

RESULT 12

US-10-156-761-2245/C
Sequence 2245, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 2245

LENGTH: 2715

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(2715)

US-10-156-761-2245

Query Match 8.4%; Score 50; DB 6; Length 2715;
Best Local Similarity 45.8%; Pred. No. 0.00013;
Matches 133; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 8 GCCTGTTGAGCGCGGCGCCGAGCGGACGACCTGTGACGATCTGCGCCGTTAGCAGGT 67
DB 1012 GCTCTTGGCGTACGCTTGAAGAGCGGACCTGCTCTGTGACGCGCGGTGACGCGCA 953
QY 68 GTGCTTGAAGCGCGTGTGCGGCGGACGCGCAGAACTGCGACGCGAAGGCGCACTCCGCGCT 127
DB 952 GGTACTTCAAGGATGTGTGTGATGCGATGCGGAAATGCGCGGCGGTGAGCGGAATCCGGGG 893
QY 128 GAGACGAGACGACGAGCGGTGTAGCGGTGCGGACGAGATCCGACCGGTGTGATGAAT 187
DB 892 ACATGTGTCGATGTGTGCGCGGCGGTGCGAGGCTGTGCGCGCACACCTTGCCTGAG 833
QY 188 TGGCGTGCACGAGCTGCAAGAGAACTTTGTGTGCTTCCCTGCGGTGCGCATGCGGCT 247
DB 832 ACTCGACGAACTTGGCCGACGACCGCGTGTGCGACGATCTGCGTGTGATGAGACGCA 773
QY 248 CGGCGAGGTTGACGAGAGGTGTGATGCTTGCCTGAGTGTGCTTCAAGCGGTGTGCGCGAGT 307
DB 772 GGTCCGATGCGGTGTGCGCGGCGGTGAGCTTCAAGCGGTGTGAGAGCGGAGCGCGG 713
QY 308 CTTTGCGCTGCCGATTAAGCTTGTGGAAGATGTGCGCTTGTGCGCGTTCGCGCGAGCGC 367

```
Db      712 GGATGACATCGAAGCCGGCTGCGGAGCATCGGCGCTCGATGCAACCGAGCC 653
QY      368 GCTCCGCGAGCTCGAGA 385
Db      652 CCCAGCCGACGACACCGA 635

RESULT 13
US-10-915-740A-17
; Sequence 17, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelein, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Maignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarfello, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Piazza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT FILING DATE: 2004-08-11
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US/10/915,740A
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 60872
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-17

Query Match      8.4%; Score 50; DB 9; Length 60872;
Best Local Similarity 54.3%; Pred. No. 0.00017;
Matches 101; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY      36 CACCTGTCGACGATCTGCGCCGGTAGACAGGTGCTTGAGAGCCGTCGCGGCCGAGCC 95
Db      8452 CAATTGCCCGATGATTTCCGCGAGCAGTCAAAATTCGCTTGAAGCCCTCCGCGCTGCA 8511
QY      96 GCAAGACTGCGACCGGAGGCGCACCCCGCTGAGACGAGCGAGGGGTGTAGCCGTC 155
Db      8512 ACAAATGTACTTCCAAAGCGCACCGCACTGTGAGAAATGCAAGAGCGTCCCGCATC 8571
QY      156 GTGGCGACGATCCGCGACCGTCTCGATGAAATTTGCGTCGACACGCTGGAACGAGACTT 215
Db      8572 CGATTCCGGGATGAAGACGCTTCCAGCCCGTTCGCCCGTACCGACATCCAAAGCCATT 8631
QY      216 TGTGCT 221
Db      8632 TCGAGT 8637

RESULT 14
US-10-915-740A-1068/c
; Sequence 1068, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelein, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Maignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarfello, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Piazza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT FILING DATE: 2004-08-11
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1068
; LENGTH: 2242716
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-1068

Query Match      8.4%; Score 50; DB 9; Length 2242716;
Best Local Similarity 54.3%; Pred. No. 0.00024;
Matches 101; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
```

```
QY      36 CACCTGTCGACGATCTGCGCCGGTAGACAGGTGCTTGAGAGCCGTCGCGGCCGAGCC 95
Db      1296347 CAATTGCCCGATGATTTCCGCGAGCAGTCAAAATTCGCTTGAAGCCCTCCGCGCTGCA 1296288
QY      96 GCAAGACTGCGACCGGAGGCGCACCCCGCTGAGACGAGCGAGGGGTGTAGCCGTC 155
Db      1296287 ACAAATGTACTTCCAAAGCGCACCGCACTGTGAGAAATGCAAGAGCGTCCCGCATC 1296228
QY      156 GTGGCGACGATCCGCGACCGTCTCGATGAAATTTGCGTCGACACGCTGGAACGAGACTT 215
Db      1296227 CGATTCCGGGATGAAGACGCTTCCAGCCCGTTCGCCCGTACCGACATCCAAAGCCATT 1296168
QY      216 TGTGCT 221
Db      1296167 TCGAGT 1296162

RESULT 15
US-10-472-928-1455/c
; Sequence 1455, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926MO
; CURRENT FILING DATE: 2003-09-26
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: GB-0107658.7
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqMan99, version 1.03
; SEQ ID NO 1455
; LENGTH: 1063
; TYPE: DNA
```

ORGANISM: Streptococcus pneumoniae
US-10-472-928-1455

Query Match 8.2%; Score 48.8; DB 8; Length 1083;
Best Local Similarity 53.7%; Pred. No. 0.00027;
Matches 101; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY	28	AGCGCCAGCAGCTGCTGCAAGATCTGCCCCGCTAGAGTGTCGCTTGAGGCGCGTCCGG	87
DB	422	ACGAGCATGATTTGCGCTACGATTTCACTTATGAGTCACTTCTTAAATCAAA	363
QY	88	CCCGACGCGGAGAACTGGCAAGCGAGGCGCACCCGCTGAGACGAGACGAGCGGTG	147
DB	362	CCAGAGGCAAGAGGTACACCGATATTAAGCCGACTGAGTGTCAACAGACAGAT	303
QY	148	TAGCCGTCGTGGGAGCGGATCCGCACTGCTCGATGAAATTGCCGTGACCAAGCTCGAC	207
DB	302	AAACCAATAGTTGAGCGCATGAGTACGCTCAATTAAATACATACGTCGGGCAATTCAAG	243
QY	208	AGGAAGTT 215	
DB	242	AGATATT 235	

Search completed: April 7, 2006, 04:35:58
Job time : 435.403 secs

This Page Blank (uspto)

CC under conditions in which the cell expresses a product of a gene encoded
CC by the nucleic acid under conditions in which the unmodified polyketide
CC is present, and producing the modified polyketide. The cell produces
CC megasamine, and can attach megasamine to a polyketide, where the cell, it
CC its naturally occurring non-recombinant state cannot produce megasamine.
CC The present sequence contains downstream megalomycin modification enzyme
CC genes.

XX Sequence 17596 BP; 2325 A; 5966 C; 6604 G; 2701 T; 0 U; 0 Other;

Query Match 100.0%; Score 597; DB 12; Length 17596;

Best Local Similarity 100.0%; Pred. No. 1.7e-132;

Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AACACGAGCTGTGAGCGCCGAGCCGAGCCGAGCAGCTGTGAGCAGATTCGCCCGGT 60
DB 17000 AACACGAGCTGTGAGCGCCGAGCCGAGCCGAGCAGCTGTGAGCAGATTCGCCCGGT 17059
QY 61 AGCAGGTGTGCTTGAAGGCGGTGCGGGCCGAGCGGAGAACTGGACAGCGAGCGCAC 120
DB 17060 AGCAGGTGTGCTTGAAGGCGGTGCGGGCCGAGCGGAGAACTGGACAGCGAGCGCAC 17119
QY 121 CCCGCTGAGAGCAGACGACGAGCGGTGTAGCCGTGTGGCGACGCGATCCGACCGTCTCG 180
DB 17120 CCCGCTGAGAGCAGACGACGAGCGGTGTAGCCGTGTGGCGACGCGATCCGACCGTCTCG 17179
QY 181 ATGAATTTGCGGTGAGCAGCTCCGAACTTTTGTCTGTGCTTCCCTTGGTGTGGA 240
DB 17180 ATGAATTTGCGGTGAGCAGCTCCGAACTTTTGTCTGTGCTTCCCTTGGTGTGGA 17239
QY 241 CTGGGCTCGGAGAGGTGTGAGAGAGGTGTGAGTTGCGGTGAGCTTCAAGCGTGTG 300
DB 17240 CTGGGCTCGGAGAGGTGTGAGAGAGGTGTGAGTTGCGGTGAGCTTCAAGCGTGTG 17299
QY 301 GCCAGTCTTTCGCTGCGCGATTAAGCTTGTGAGATGTGCGGTCTTTCGCGCG 360
DB 17300 GCCAGTCTTTCGCTGCGCGATTAAGCTTGTGAGATGTGCGGTCTTTCGCGCG 17359
QY 361 CCGAGCGGCTTCGCGAGCTTCGAGAGACAGCAGTGTGAAGACCGAGCGCGAGCGGTGT 420
DB 17360 CCGAGCGGCTTCGCGAGCTTCGAGAGACAGCAGTGTGAAGACCGAGCGCGAGCGGTGT 17419
QY 421 CGTGCGGATGGGTGAGCCACGACGCTGGGGCGAAGCTGACATGTGACCACTATCA 480
DB 17420 CGTGCGGATGGGTGAGCCACGACGCTGGGGCGAAGCTGACATGTGACCACTATCA 17479
QY 481 CGGTGCAAGAGACGTCAATTGCTCAAGTGAACCAAGAGGCTTGACATGAGCATGCT 540
DB 17480 CGGTGCAAGAGACGTCAATTGCTCAAGTGAACCAAGAGGCTTGACATGAGCATGCT 17539
QY 541 CTCGTGTCTTGGCCATATAGCGGTGTGAGCTGCGCAATTGACGAACCGCGAGCGGCGC 597
DB 17540 CTCGTGTCTTGGCCATATAGCGGTGTGAGCTGCGCAATTGACGAACCGCGAGCGGCGC 17596
```

RESULT 2
AAI99682_31
Continuation (32 of 45) of AAI99682 from base 3100001 (Mycobacterium tuberculosis strain
WP Sequence Split into 45 fragments LOCUS AAI99682 Accession AAI99682

Fragment Name	Begin	End
WP AAI99682_00	1	110000
WP AAI99682_01	100001	210000
WP AAI99682_02	200001	310000
WP AAI99682_03	300001	410000
WP AAI99682_04	400001	510000
WP AAI99682_05	500001	610000
WP AAI99682_06	600001	710000
WP AAI99682_07	700001	810000
WP AAI99682_08	800001	910000
WP AAI99682_09	900001	1010000
WP AAI99682_10	1000001	1110000
WP AAI99682_11	1100001	1210000
WP AAI99682_12	1200001	1310000

```
WP AAI99682_13 130001 1410000
WP AAI99682_14 140001 1510000
WP AAI99682_15 150001 1610000
WP AAI99682_16 160001 1710000
WP AAI99682_17 170001 1810000
WP AAI99682_18 180001 1910000
WP AAI99682_19 190001 2010000
WP AAI99682_20 200001 2110000
WP AAI99682_21 210001 2210000
WP AAI99682_22 220001 2310000
WP AAI99682_23 230001 2410000
WP AAI99682_24 240001 2510000
WP AAI99682_25 250001 2610000
WP AAI99682_26 260001 2710000
WP AAI99682_27 270001 2810000
WP AAI99682_28 280001 2910000
WP AAI99682_29 290001 3010000
WP AAI99682_30 300001 3110000
WP AAI99682_31 310001 3210000
WP AAI99682_32 320001 3310000
WP AAI99682_33 330001 3410000
WP AAI99682_34 340001 3510000
WP AAI99682_35 350001 3610000
WP AAI99682_36 360001 3710000
WP AAI99682_37 370001 3810000
WP AAI99682_38 380001 3910000
WP AAI99682_39 390001 4010000
WP AAI99682_40 400001 4110000
WP AAI99682_41 410001 4210000
WP AAI99682_42 420001 4310000
WP AAI99682_43 430001 4410000
WP AAI99682_44 440001 441529
```

Query Match 13.0%; Score 77.8; DB 4; Length 110000;
Best Local Similarity 61.7%; Pred. No. 1.4e-08;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
QY 26 CGAGCGCGAGCAGCTGTGCAAGATCTGCGCCGTTAGCAGGTGTGCTTGAAGCGGTGCG 85
DB 90207 CGCGGCGAGCAGCTGTGCAAGATCTGCGCCGTTAGCAGGTGTGCTTGAAGCGGTGCG 90266
QY 86 GCGCCGAGCGCGAGACTGTGCGAGCGGAGAGCGCACCCCGCTTGAGACGAGCGAGCGG 145
DB 90267 GCGCCGAGCGCGAGACTGTGCGAGCGGAGAGCGCACCCCGCTTGAGAGCGAGCGG 90326
QY 146 TGTAGCGGTGTGCGAGCGATTCGCAACCGCTGATGAAATTGCGCGACGACCTGGA 205
DB 90327 TGTAGCGGTGTGCGAGCGATTCGCAACCGCTGATGAAATTGCGCGACGACCTGGA 90386
QY 206 ACAGGAATTTGTGCTGCGC 226
DB 90387 ACAGGAATTTGTGCTGCGC 90407
```

RESULT 3
AAI99683_31

Continuation (32 of 44) of AAI99683 from base 3100001 (Mycobacterium tuberculosis strain
WP Sequence Split into 44 fragments LOCUS AAI99683 Accession AAI99683

Fragment Name	Begin	End
WP AAI99683_00	1	110000
WP AAI99683_01	100001	210000
WP AAI99683_02	200001	310000
WP AAI99683_03	300001	410000
WP AAI99683_04	400001	510000
WP AAI99683_05	500001	610000
WP AAI99683_06	600001	710000
WP AAI99683_07	700001	810000
WP AAI99683_08	800001	910000
WP AAI99683_09	900001	1010000
WP AAI99683_10	1000001	1110000
WP AAI99683_11	1100001	1210000
WP AAI99683_12	1200001	1310000
WP AAI99683_13	1300001	1410000

WP AA19683_14 1400001 1510000
XX AA19683_15 1500001 1610000
WP AA19683_16 1600001 1710000
WP AA19683_17 1700001 1810000
WP AA19683_18 1800001 1910000
WP AA19683_19 1900001 2010000
WP AA19683_20 2000001 2110000
WP AA19683_21 2100001 2210000
WP AA19683_22 2200001 2310000
WP AA19683_23 2300001 2410000
WP AA19683_24 2400001 2510000
WP AA19683_25 2500001 2610000
WP AA19683_26 2600001 2710000
WP AA19683_27 2700001 2810000
WP AA19683_28 2800001 2910000
WP AA19683_29 2900001 3010000
WP AA19683_30 3000001 3110000
WP AA19683_31 3100001 3210000
WP AA19683_32 3200001 3310000
WP AA19683_33 3300001 3410000
WP AA19683_34 3400001 3510000
WP AA19683_35 3500001 3610000
WP AA19683_36 3600001 3710000
WP AA19683_37 3700001 3810000
WP AA19683_38 3800001 3910000
WP AA19683_39 3900001 4010000
WP AA19683_40 4000001 4110000
WP AA19683_41 4100001 4210000
WP AA19683_42 4200001 4310000
WP AA19683_43 4300001 4403765

Query Match 13.0%; Score 77.8; DB 4; Length 110000;
Best Local Similarity 61.7%; Pred. No. 1.4e-08;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 26 CGAGCGCGCAGCACTCTGCAAGATCTCCGCCGTAGCAGTGTCTGAGGCCCTGCC 85
DB 84316 CCGCGGACGACGACTCTGAGAGATCTCCGCCGTGATGTTGCGAATCCGCCCT 84375
QY 86 GCGCCGACGCGCAAGACTGCGCAGCGCAAGCGGCAACCCCGCTTGAAGACGACGCGG 145
DB 84376 GCGCAGTGGCAGCAAGACGCGGCAAGCCATCCCGACCGGCTGCGGAATGCAAGCCG 84435
QY 146 TGTAGCGCTGTGGCGCAGGATCCGACCGCTCTGATGAATTTGCGTGAACGAGTCCA 205
DB 84436 TGTGGCGCGCGGATAGCGCATTCAGATTCGAATGTTACCGTGAACGCGCCGCC 84495
QY 206 ACAGAACTTTGCTGTGCGC 226
DB 84496 ACAACGCTTTTCGAGTCTGCGC 84516

RESULT 4
ACL73225/c
ID ACL73225 standard; DNA; 1080 BP.

XX AC ACL73225;
XX DT 02-JUN-2005 (first entry)
XX DE M. xanthus gene sequence, seq id 9688.
XX KW Transgenic plant; DNA replication; gene regulation; gene expression;
XX KM gene; db.
XX OS Myxococcus xanthus.
XX PN US6833447-B1.
XX PD 21-DEC-2004.
XX PF 10-JUL-2001; 2001US-00902540.
XX

PR 10-JUL-2000; 2000US-0217883P.
XX (MONS) MONSANTO TECHNOLOGY LLC.
PA Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX WPI; 2005-028716/03.
DR

XX New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.

Example 2; SEQ ID NO 9688; 25pp; English.

CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC set of about 7842 genes or partial genes from the genome of the bacterium
CC Myxococcus xanthus. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from USPRO

Sequence 1080 BP; 173 A; 403 C; 343 G; 161 T; 0 U; 0 Other;

Query Match 13.0%; Score 77.6; DB 14; Length 1080;
Best Local Similarity 61.3%; Pred. No. 8.3e-09;
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 34 AGCACTGTGCAACATCTCGCCCGGTACAGGTGTGCTTAGAGCCGTGCGGCCGAC 93
DB 431 AGTATGTGCCACGATTTGCGCGGCGTCAAGTGGCGTGAAGGCCCATCTGCCGGTG 372
QY 94 GCGCAGAACTGCGACGCGGACGCGCAACCCCGCTTGAAGACGACGCGGTGTGCGG 153
DB 371 GCGCAGAAAGACGACGCGGACGCGGCGCGCTGTGCTGATGCAACCGTGGCGCGG 312
QY 154 TGTGTGCGAGGATCGCAGCGCTCTGATGAATTTGCGTGAACGAGTCTGAACGAGAAC 213
DB 311 CCTTGAACCGATGACGACGCTTCAATGATGCTGCCCTGTCCAGTGGCAGCAGCAGC 252
QY 214 TTTGTGCTGTGCTTCCCTGTG 237
DB 251 TTGTGCGTGAACCGTGTGCTG 228

RESULT 5
ACL64636/c
ID ACL64636 standard; DNA; 13751 BP.

XX AC ACL64636;
XX DT 02-JUN-2005 (first entry)
XX DE M. xanthus DNA fragment, seq id 1099.
XX KW Transgenic plant; DNA replication; gene regulation; gene expression; db.
XX KM Myxococcus xanthus.
XX OS US6833447-B1.
XX PN 21-DEC-2004.
XX PD 10-JUL-2001; 2001US-00902540.
XX PF 10-JUL-2000; 2000US-0217883P.
XX (MONS) MONSANTO TECHNOLOGY LLC.

XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
PI WPI; 2005-028716/03.
XX
XX
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
XX Example 1; SEQ ID NO 1099; 25bp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium *Myxococcus xanthus*. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 13751 BP; 2088 A; 5011 C; 4651 G; 2001 T; 0 U; 0 Other;
Query Match 13.0%; Score 77.6; DB 14; Length 13751;
Best Local Similarity 61.3%; Pred. No. 1.2e-08;
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 34 AGCAGCTGCTGACGATCTGCGCGGTAGAGTGTGCTTGAAGCGCTGCGGCGGAC 93
DB 5590 AGTATCTGCCCCAGATTTCCCGGCGCTCAGGTGGCTGAGAGCCCATCTGCCCCGAT 5531
QY 94 GCGCAGAACTGCGACGCGAAGCGCGACCCCGCTGAGAGCAGACGCGAGCGGTAGCCG 153
DB 5530 GCGCAGAAAGACGACGCCATGCGGCGAGCGGCTGCGGTGATGACACCGTGGCGCG 5471
QY 154 TCGTGGCAGCGAATCCGCGACCGTCTCATGAAATTCGCTGACCACTCGAAGCAGAAC 213
DB 5470 CCGTTCGACCGCATCAGCAGCGTTCATGATGTGCGCGGTGCTCAGTCCAGCAGCAGC 5411
QY 214 TTGTGCTGTGCGCTTCCCTGTGATG 237
DB 5410 TTGTGCGTGAAGCCGTGCTGCTG 5387

RESULT 6
ACL6587/c
ID ACL6587 standard; DNA; 792 BP.
XX
XX ACL6587;
AC
XX
XX 02-JUN-2005 (first entry)
DT
XX
XX M. xanthus gene sequence, seq id 3050.
DE
XX
XX Transgenic plant; DNA replication; gene regulation; gene expression;
KW gene; ds.
XX
XX *Myxococcus xanthus*.
OS
XX
XX US6833447-B1.
PN
XX
XX 21-DEC-2004.
PD
XX
XX 10-JUL-2001; 2001US-00902540.
PF
XX
XX 10-JUL-2000; 2000US-0217883P.
PR
XX
XX (MONS) MONSANTO TECHNOLOGY LLC.
PA
XX

PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX WPI; 2005-028716/03.
XX
XX
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
XX Example 2; SEQ ID NO 3050; 25bp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
CC set of about 7842 genes or partial genes from the genome of the bacterium
CC *Myxococcus xanthus*. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from USPTO
XX
SQ Sequence 792 BP; 148 A; 267 C; 257 G; 120 T; 0 U; 0 Other;
Query Match 12.6%; Score 75; DB 14; Length 792;
Best Local Similarity 58.9%; Pred. No. 3.3e-08;
Matches 129; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 3 CACGAGCTTGTGACGCGCGGCCCGGACCGCAGACCTGCTGCACGATCTGCCCGGTAG 62
DB 459 CTCCTTTTTCGAAACCTCGCGGTTCACCGGTGACCTGGAGCCACATCTCGTGGAGGT 400
QY 63 CAGGTGTGCTTGAAGCGCGGTGCGCGGCGGACGCGCAGAACTGCGACCGGAGCGCACCC 122
DB 399 CAGGTGTGCTTGAAGCGCGGTGCGCGGCGGACGCGCAGAACTGCGACCGGAGCGCACCC 340
QY 123 CGCTGAGACGAGACGCGAGCGGTGTAGCCGTGCTGCGAAGATCCGACCGTCTGAT 182
DB 339 CACCTGTGTGAGACGCGAAGCGTCTGCGGTCTCGGTGGGATGTAGACGATTCGAT 280
QY 183 GAAATTCGCTGACGACGCTGGAACGAACTTTGTGT 221
DB 279 GTACGCGCGGTCCGCGCTTTCACCGGTACTTGTATGT 241

RESULT 7
ACL64330
ID ACL64330 standard; DNA; 6713 BP.
XX
XX ACL64330;
AC
XX
XX 02-JUN-2005 (first entry)
DT
XX
XX M. xanthus DNA fragment, seq id 793.
DE
XX
XX Transgenic plant; DNA replication; gene regulation; gene expression; ds.
KW
XX
XX *Myxococcus xanthus*.
OS
XX
XX US6833447-B1.
PN
XX
XX 21-DEC-2004.
PD
XX
XX 10-JUL-2001; 2001US-00902540.
PF
XX
XX 10-JUL-2000; 2000US-0217883P.
PR
XX
XX (MONS) MONSANTO TECHNOLOGY LLC.
PA
XX
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
PI WPI; 2005-028716/03.
XX

XX New substantially purified *Mycrococcus xanthus* nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 1; SEQ ID NO 793; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contigs and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium *Mycrococcus xanthus*. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 6713 BP; 984 A; 2232 C; 2234 G; 1263 T; 0 U; 0 Other;
Query Match 12.6%; Score 75; DB 14; Length 6713;
Best Local Similarity 58.9%; Pred. No. 4.4e-08;
Matches 129; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 3 CAGAGCGCTTGAACGCGGCGCCGAGCGCCAGACACTCTGTCAGATTCGCGCCGTTAG 62
Db 336 CTGCTGTTCTTGAACCTCGCGGTTTACCGCGTGAACCTCGGCGCCAGATTCCTGCGCGGGT 395
QY 63 CAGGTGCGCTTGAAGCGCGCTCGCGCGCCGAGCGCCAGACACTCTGTCAGATTCGCGCCG 122
Db 396 CAGGTGCGCTTGAAGCGCGCTCGCGCGCCGAGCGCCAGACACTCTGTCAGATTCGCGCCG 455
QY 123 CGCTTGAAGACGAGACGCGCGGTGTAGCGCTGTCGCGAGATTCGCGACCTCTTCGAT 182
Db 456 CACCTGCGTGAACGAGACGCGCGGTGTAGCGCTGTCGCGAGATTCGCGACCTCTTCGAT 515
QY 183 GAATTGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCT 221
Db 516 GTAGCGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCTTGAACGCGCT 554

RESULT 8
ID ABQ81849 standard; DNA; 349980 BP.
XX
AC ABQ81849;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1105.
XX
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KM antidiarrhetic; antibacterial; inhibitor of *Salmonella*; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KM rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
XX Bifidobacterium longum.
OS Synthetic.
OS
PN EP1227152-A1.
XX
PD 31-UTL-2002.
XX
PF 30-JAN-2001; 2001EP-00102050.
XX
PR 30-JAN-2001; 2001EP-00102050.
XX
XX (NEST) SOC PROD NESTLE SA.
PA
XX
DR WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT a probe or primer for detecting and/or identifying Bifidobacterium longum
PT in a biological sample.
XX
PS Disclosure; SEQ ID NO 1105; 80pp; English.
XX
CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC fusion protein, comprising a sequence selected from 1097 sequences given
CC in ABQ65258 to ABQ65354 ligated in frame to a polynucleotide encoding a
CC heterologous polypeptide. (I) has antidiarrhetic and antibacterial
CC activities, and can be used as an inhibitor of *Salmonella*. (I) (which is
CC a probe) is useful for the detection and/or identification of
CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-creams, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
XX
SQ Sequence 349980 BP; 68046 A; 106490 C; 106389 G; 69055 T; 0 U; 0 Other;
Query Match 12.4%; Score 73.8; DB 6; Length 349980;
Best Local Similarity 60.0%; Pred. No. 1.5e-07;
Matches 123; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 26 CGAGCGCCGACGACCTCTGTCAGATTCGCGCGGTGTAGCGCTTGAAGCGCGTTCGCG 85
Db 85216 CGGCGACGACGACCTCTGTCAGATTCGCGCGGTGTAGCGCTTGAAGCGCGTTCGCG 85157
QY 86 GCGCCGACGCGGACGACCTCTGTCAGATTCGCGCGGTGTAGCGCTTGAAGCGCGTTCGCG 145
Db 85156 TGCGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGG 85097
QY 146 TGTAGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGG 205
Db 85096 TGTAGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGGACGCGGTGTAGCGG 85037
QY 206 ACAGGAACTTGTGCTGCTGCTTCC 230
Db 85036 ACAGGAACTTGTGCTGCTGCTTCC 85012

RESULT 9
ID ABD07607 standard; DNA; 456 BP.
XX
AC ABD07607;
XX
DT 29-UTL-2004 (first entry)
XX
DE *Pseudomonas aeruginosa* polynucleotide #6211.
XX
XX Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
KM antibacterial.
XX
XX *Pseudomonas aeruginosa*.
OS
XX
PN US6551795-B1.

XX PD 22-APR-2003.
XX PF 18-FEB-1999; 99US-00252991.
XX PR 18-FEB-1998; 98US-0074788P.
XX PR 27-JUL-1998; 98US-0094190P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Rubinfeld MJ, Nolling J, Deloughery C, Bush D;
XX DR WPI; 2003-615309/58.
XX DR P-PSDB; ABO74036.
XX PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 6211; 455bp; English.
XX CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the
XX CC polymucleotides encoding them. The sequences are useful in diagnosis and
XX CC therapy of pathological conditions, as molecular targets for diagnostics,
XX CC prophylaxis and treatment of pathological conditions resulting from a
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,
XX CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
XX CC effective antibacterial targets, as targets for antibacterial drugs,
XX CC including anti-*P. aeruginosa* drugs, as templates for recombinant
XX CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
XX CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
XX CC infection, and in detection of *P. aeruginosa* sequences or other sequences
XX CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-
XX CC ABD17967 represent *P. aeruginosa* polymucleotides of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format from USPTO at
XX CC seqdata.uspto.gov/sequence.html
XX SQ Sequence 456 BP; 82 A; 142 C; 147 G; 85 T; 0 U; 0 Other;
Query Match 10.9%; Score 64.8; DB 11; Length 456;
Best Local Similarity 60.0%; Pred. No. 8.3e-06;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 36 CACCTGTGACGATCTGCGCCGGTAGACAGGTGTGCTTGAGAGCCGTCGCGCCGACGC 95
DB 144 CACCTGGCCGATCACTCGGCGGGGTAGAGTGTGCTTGTAACCTCTGCTGCGGTGGA 203
QY 96 GCAGAACTGCGACCGGAAAGGCGCACCCGCTTGAGACGACGACGGCGGTGTAGCCGTC 155
DB 204 GCAGAACTGCGAATCCAGGGGCGACGCCGCTGGGAGGACACGACAGGGTGGCGCTCC 263
QY 156 GTGGCGACGATCCGCGACCGCTTGATGAATTTGCCCTGACACGACTGGAACAGGAATT 215
DB 264 GCCCTGGGGAGTGTACACGCTTCGACGCAACTGCCGAGCGCACCCGAGACCACTT 323

RESULT 10
ABD07733/c
ID ABD07733 standard; DNA; 855 BP.
XX AC ABD07733;
XX XX 29-JUL-2004 (first entry)
XX DE *Pseudomonas aeruginosa* polymucleotide #6337.
XX XX Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
XX KW antibacterial.
XX OS *Pseudomonas aeruginosa*.
XX XX US6551795-B1.

XX PD 22-APR-2003.
XX PF 18-FEB-1999; 99US-00252991.
XX PR 18-FEB-1998; 98US-0074788P.
XX PR 27-JUL-1998; 98US-0094190P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Rubinfeld MJ, Nolling J, Deloughery C, Bush D;
XX DR WPI; 2003-615309/58.
XX DR P-PSDB; ABO74162.
XX PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 6337; 455bp; English.
XX CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the
XX CC polymucleotides encoding them. The sequences are useful in diagnosis and
XX CC therapy of pathological conditions, as molecular targets for diagnostics,
XX CC prophylaxis and treatment of pathological conditions resulting from a
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,
XX CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
XX CC effective antibacterial targets, as targets for antibacterial drugs,
XX CC including anti-*P. aeruginosa* drugs, as templates for recombinant
XX CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
XX CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
XX CC infection, and in detection of *P. aeruginosa* sequences or other sequences
XX CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-
XX CC ABD17967 represent *P. aeruginosa* polymucleotides of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format from USPTO at
XX CC seqdata.uspto.gov/sequence.html
XX SQ Sequence 855 BP; 168 A; 270 C; 263 G; 154 T; 0 U; 0 Other;
Query Match 10.9%; Score 64.8; DB 11; Length 855;
Best Local Similarity 60.0%; Pred. No. 9.1e-06;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 36 CACCTGTGACGATCTGCGCCGGTAGACAGGTGTGCTTGAGAGCCGTCGCGCCGACGC 95
DB 436 CACCTGGCCGATCACTCGGCGGGGTAGAGTGTGCTTGTAACCTCTGCTGCGGTGGA 377
QY 96 GCAGAACTGCGACCGGAAAGGCGCACCCGCTTGAGACGACGACGGCGGTGTAGCCGTC 155
DB 376 GCAGAACTGCGAATCCAGGGGCGACGCCGCTGGGAGGACACGACAGGGTGGCGCTCC 317
QY 156 GTGGCGACGATCCGCGACCGCTTGATGAATTTGCCCTGACACGACTGGAACAGGAATT 215
DB 316 GCCCTGGGGAGTGTACACGCTTCGACGCAACTGCCGAGCGCACCCGAGACCACTT 257

RESULT 11
ABD07815/c
ID ABD07815 standard; DNA; 1164 BP.
XX AC ABD07815;
XX XX 29-JUL-2004 (first entry)
XX DE *Pseudomonas aeruginosa* polymucleotide #6419.
XX XX Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
XX KW antibacterial.
XX OS *Pseudomonas aeruginosa*.
XX XX US6551795-B1.

XX 22-APR-2003.
PD 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO74244.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 6419; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01357-
CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1164 BP; 234 A; 361 C; 356 G; 213 T; 0 U; 0 Other;
Query Match 10.9%; Score 64.8; DB 11; Length 1164;
Best Local Similarity 60.0%; Pred. No. 9.5e-06;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 36 CACCTGTCGACGATCTCGCCGCGTAGCAGTGTCTGAGGCGCGGCCGACGC 95
DB 456 CACCTGCGCGATCACTCGCGCGGTGAGTGTCTGTGAAACCTGCTTGCCTGCGTGA 397
QY 96 GCAGAACTGGCAGCGGAGGCGGACCCGCGCTGAGACGAGCGAGCGGTGTAGCCGTC 155
DB 396 GCAGAACTGCAATTCAGGGCGGAGCCCGCTGGAGAGACGACAGGGGTGCGGCTTC 337
QY 156 GTGGCGACGAGATCCGACCGTCTCGATGAAATTGCCGTGACGAGCTCGAAGCACTT 215
DB 336 GCCCTGGGGGATTAACGCGTCTCGACGAACTGCGGAGCGCAACCGGACACCACTT 277
RESULT 12
ABD07529
ID ABD07529 standard; DNA; 1764 BP.
XX
AC ABD07529;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #6133.
XX
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.

XX 22-APR-2003.
PD 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO73958.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 6133; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01357-
CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1764 BP; 308 A; 544 C; 557 G; 355 T; 0 U; 0 Other;
Query Match 10.9%; Score 64.8; DB 11; Length 1764;
Best Local Similarity 60.0%; Pred. No. 1e-05;
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 36 CACCTGTCGACGATCTCGCCGCGTAGCAGTGTCTGAGGCGCGGCCGACGC 95
DB 1525 CACCTGCGCGATCACTCGCGCGGTGAGTGTCTGTGAAACCTGCTTGCCTGCGTGA 1584
QY 96 GCAGAACTGGCAGCGGAGGCGGACCCGCGCTGAGACGAGCGAGCGGTGTAGCCGTC 155
DB 1585 GCAGAACTGCAATTCAGGGCGGAGCCCGCTGGAGAGACGACAGGGGTGCGGCTTC 1644
QY 156 GTGGCGACGAGATCCGACCGTCTCGATGAAATTGCCGTGACGAGCTCGAAGCACTT 215
DB 1645 GCCCTGGGGGATTAACGCGTCTCGACGAACTGCGGAGCGCAACCGGACACCACTT 1704
RESULT 13
ACL35016
ID ACL35016 standard; cDNA; 2000 BP.
XX
AC ACL35016;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:13579.
XX
XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.

XX 30-JAN-2003.
PD
XX
XX 21-JUN-2002; 2002WO-US019668.
PF
XX 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Kieps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX WPI; 2003-248011/24.
DR
XX New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
XX Claim 48; SEQ ID NO 13579; 89pp; English.
PS
XX The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC molecule to an abiotic stress responsive polynucleotide. The nucleic acid
CC and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 342 A; 599 C; 666 G; 389 T; 0 U; 4 Other;
Query Match 10.7%; Score 63.8; DB 11; Length 2000;
Best Local Similarity 55.1%; Pred. No. 1.8e-05;
Matches 125; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 11 TGTGAGCGCGGCGCCGAGCCAGCACTGTCGACATCTCCGCCGTGAGCGTTC 70
DB 711 TGCCAGGTGCGTGCAGCAACCACTGCGCATATCTCGCGGTGCTCAGTTGC 770
QY 71 GCTTGAAGCGCTGCGGCGCCGAGCGAGAACTGGCAGCGAGCGCACCCCGCTGAG 130
DB 771 GATTGAAGCCTTGAGGTGCGGTGAGGAGAAAGTGCAGTTCAAGGCCGACACCTGCG 830
QY 131 ACGAGACCGCAGCGGTGTAAGCGCTGTCGCAAGATCCGACCGTCTCGATGAATTC 190
DB 831 ACGACACCGCAGCGGTGCGGCGGTCTTGTCCGGATGTACACGGTCTCGATGCGCTTCT 890
QY 191 CGTGCACCGCTCGAAGCAAGAACTTGTGCTGCGCTTCCCGTGTG 237
DB 891 TGCCATCCAGCGCCATCGCCAGCAGCCACTTGTGCGTCCGTGCGG 937

RESULT 14
AAH67179/c
ID AAH67179 standard; DNA; 1098 BP.

XX AAH67179;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum coding sequence fragment SEQ ID NO: 2214.
DE
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.

XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
FN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-00127688.
PF
XX 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
DR P-PSDB; AAG91960.
DR
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 8; SEQ ID NO 2214; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and analyzing
CC the expression profile or expression pattern of a gene derived from
CC Corynebacterium bacterium, and identifying a homolog of a gene derived from
CC Corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX
SQ Sequence 1098 BP; 221 A; 295 C; 304 G; 278 T; 0 U; 0 Other;
Query Match 10.5%; Score 62.6; DB 5; Length 1098;
Best Local Similarity 59.1%; Pred. No. 3.1e-05;
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 35 GCACCTGCTGACGATCTCGCCCGGTGAGCGAGTGTCCCTGAGGCGCGGCCGACG 94
DB 466 GAACCTGATCCAGCATCTCACCGATTGAAGGTGACGCTCCAAACCGCCTGACCAAGTTG 407
QY 95 CGCAGAACTGGCAGCGGAGCGGAGCGCCGCTGAGAGAGACGAGCGGTGAGCGGT 154
DB 406 CACGAAATGGCAGCGCATGCGGACCAAGCTCGAAGAAATACAGCGCTGAGAGCAT 347
QY 155 CGTGCAGCGGATCGGACCGCTCTCGATGAATTCGCTGACCAAGCTCGAAGCAAGAACT 214
DB 346 CTGAATGCGGATGAGAAAGAGACTCAAGCAAAAGGCATCATGAGCTTCATTAACGCT 287
QY 215 T 215
DB 286 T 286

RESULT 15
ACA00654/c
ID ACA00654 standard; DNA; 1098 BP.

XX ACA00654;
AC
XX 03-JUN-2003 (first entry)
DT
XX C. glutamicum derived ORF SEQ ID 645.
DE
XX

KM Coryneform; nucleic acid array; fermentation; culture; ds.
XY

OS *Corynebacterium glutamicum*.
VY

PN DE10128510-A1.
XX

PD 19-DEC-2002.

PF 13-JUN-2001; 2001DE-01028510.
VV

PR 13-JUN-2001; 2001DE-01028510.

PA (DEGS) DEGUSSA AG.

PI Farwick M, Moeckel B, Pfefferle W, Bathe B, Hutmacher K; vv

DR WPI; 2003-279970/28.

PT New nucleic acid array useful for monitoring mRNA expression of

PT From *Corynebacterium glutamicum*.
VV

PS Claim 1; Page 269; 709pp; German.

CC This invention describes a novel
CC Combinations of substituents and/or

CC analyse *C. glutamicum*, particularly for monitoring a fermentation process

CC monitoring particularly differentiates between expression levels of

CC culture and fermentation conditions. ACA00010-ACA0218 represent C. glutamicum derived polymers described in the patent literature.

	Invention	CC BY
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		

SQ Sequence 1098 BP; 221 A; 295 C; 304 G; 278 T; 0 U; 0 Other;

Query Match 10.5%; Score 62.6; DB 8; Length 1098;

Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0

35 GCACCTGCTGCACGATCTCGCCCGTAGCAGGTGTCCCTTGAGGCCGTCGCGGCCGACG 94

Db 466 GAACCTGATCCACGATCTCACC GATTGAAGGTTACGGTCCAAACCGCCCTGACCAGTTG 407

95 CGCAGACTGGCAGCGGACCCCGCTGAGACGAGACCGAGCGGTGACCGT 154

Db 406 CACAGATGGCAGCCCATGCCGCAACCAAGCCTGCCGAGAATACACAGCGTGAGCGAT 347

155 CGTGGCGACGGATCCGCACCGTCTCGATGAATTGCCGTCGACCAGCTCGACAGGA 214

Db 346 CTGAATAGCGCATGAGAACGACTCAAGCAAAGTGCCTCATGGAGCTTCCATAACGCT 287

QY 215 T 215

Db 286 T 286

[REDACTED]

Job time : 343.371 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:24:14 ; Search time 2885.48 Seconds
(Without alignments)
11760.789 Million cell updates/sec

Title: US-10-611-442-2_COPY_17000_17596

Perfect score: 597

Sequence: 1 aacacgagcctgtgacgc.....cacgaacgcgcagcgcgccg 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBml:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	87.4	14.6	110000	1	BA000043_11
C 2	87	14.6	301332	1	AE017237
C 3	84.4	14.1	110000	1	BA000045_22
C 4	83.2	13.9	308050	1	SC0939124
C 5	81.2	13.6	207050	1	AL646063
C 6	78.4	13.1	110000	1	AP006618_42
C 7	77.8	13.0	110000	1	AE000516_31
C 8	77.8	13.0	110000	1	AE017333_17
C 9	77.8	13.0	110000	1	CP000002_17
C 10	77.8	13.0	318050	1	BX248344
C 11	77.6	13.0	348676	1	BX842581
C 12	77.6	13.0	1080	6	AR628287
C 13	77.6	13.0	13751	6	AR619704
C 14	77.2	12.9	110000	1	BA000030_32
C 15	76	12.7	10719	1	AE001946
C 16	75	12.6	792	6	AR621649
C 17	75	12.6	6713	6	AR619398
C 18	75	12.6	110000	1	BX571965_17

19	74.6	12.5	347894	1	BX640431
20	74.6	12.5	348642	1	BX640446
21	74.6	12.5	348934	1	BX640417
C 22	73.8	12.4	110000	1	AE014295_04
C 23	73.8	12.4	349980	6	AX492787
C 24	73.8	12.4	349980	6	AX53954
C 25	73.4	12.3	110000	1	CP000010_13
C 26	73.4	12.3	110000	1	CP000010_14
C 27	72.6	12.2	110000	1	CP000090_22
C 28	72.2	12.1	110000	1	BA000035_20
C 29	72	12.1	110000	1	CP000088_07
C 30	69.6	11.7	110000	1	AE017282_30
C 31	68	11.4	300217	1	AE016922
C 32	67	11.2	110000	1	AP006840_14
C 33	66.2	11.1	110000	1	CR555306_07
C 34	65.8	11.0	110000	1	AP008231_24
C 35	65.6	11.0	110000	1	AE017180_11
C 36	65	10.9	110000	1	CR931997_13
C 37	65	10.9	110000	1	CP000089_32
C 38	64.8	10.9	1993	1	PSEB1P
C 39	64.8	10.9	17148	1	AE004798
C 40	64.2	10.8	349652	1	BX569690
C 41	63.6	10.7	49742	1	AP008226_18
C 42	63.2	10.6	348408	1	BX248358
C 43	62.6	10.5	1098	6	BD164415
C 44	62.6	10.5	1098	6	AX122298
C 45	62.6	10.5	110000	1	BA000036_21

ALIGNMENTS

RESULT 1	BA000043_11/c	WPCOMENT
Sequence split into 36 fragments	LOCUS BA000043	Accession BA000043
Fragment Name	Begin	End
BA000043_00	1	110000
BA000043_01	100001	210000
BA000043_02	200001	310000
BA000043_03	300001	410000
BA000043_04	400001	510000
BA000043_05	500001	610000
BA000043_06	600001	710000
BA000043_07	700001	810000
BA000043_08	800001	910000
BA000043_09	900001	1010000
BA000043_10	1000001	1110000
BA000043_11	1100001	1210000
BA000043_12	1200001	1310000
BA000043_13	1300001	1410000
BA000043_14	1400001	1510000
BA000043_15	1500001	1610000
BA000043_16	1600001	1710000
BA000043_17	1700001	1810000
BA000043_18	1800001	1910000
BA000043_19	1900001	2010000
BA000043_20	2000001	2110000
BA000043_21	2100001	2210000
BA000043_22	2200001	2310000
BA000043_23	2300001	2410000
BA000043_24	2400001	2510000
BA000043_25	2500001	2610000
BA000043_26	2600001	2710000
BA000043_27	2700001	2810000
BA000043_28	2800001	2910000
BA000043_29	2900001	3010000
BA000043_30	3000001	3110000
BA000043_31	3100001	3210000
BA000043_32	3200001	3310000
BA000043_33	3300001	3410000
BA000043_34	3400001	3510000
BA000043_35	3500001	3544776


```
gene
CDS
    complement (6994..7788)
    /locus_tag="MAP2703c"
    complement (6994..7788)
    /locus_tag="MAP2703c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA050502.1"
    /db_xref="GI:41397160"
    /translation="MEVEIIPRLKEPLYRYIEMRLROELGAADLPRIAVICDGN
RRMADAGDYDVSYGRMGAVKIAEMLRMCOAGIEMTYVYLITSTENLQRPDELGL
IRIITDVYERICAPANRMSVRYTGDIJELGSEPARLRGAVESTPYEASHVNAVGY
GGRREIVDAVRSLISKELANGATAEELIDAVTVDAISENLVYSGQDPDLVIRTSGEQ
RLSGFLMQSAVSEMFTEAHMPAFRVDLRLALRDYSGRHRVYGR"
    7961..8734
    /locus_tag="MAP2704"
    7961..8734
    /locus_tag="MAP2704"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA050502.1"
    /db_xref="GI:41397161"
    /translation="MSSQSTAPNPEPDHEALVPGNTVEQLPFGAAQVIGKPMRGIT
HFSAMLAIVTGAITLVSVSMAASPRAGSTLIYAATAVLPVSAVYRVRKKTAA
RTRMKRLDHSMTFVPIAGSYTPPARLVMPQNTGVVLCIWMGLGALGAIKVCMPPTAP
RM/GVLYLLGLGVAIWMVAPITLIHQGVANAMVLAAGVLYSVGGIFYALRMDDPMS
TGYHEFFHALTIAALITVYIYIWMFAVFTGDHAWLATP"
    complement (8744..9127)
    /locus_tag="MAP2705c"
    complement (8744..9127)
    /locus_tag="MAP2705c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA050502.1"
    /db_xref="GI:41397162"
    /translation="MPNDAKTDAIRNTVRYIELVAKGADDLVELYADATVEDPVG
GVHIGRQAIHGFSYALDNILQRECELTILRVAGNEAAFFRLITVAGBHRMYLEPIDV
WYEDNGKVAAMKAYMSPDIDVSRP"
    complement (9149..11164)
    /locus_tag="MAP2706c"
    complement (9149..11164)
    /locus_tag="MAP2706c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA050503.1"
    /db_xref="GI:41397163"
    /translation="MSPADPSATNTLGRTSPYLRQHADNPVYMQWTPQALADAAR
DVPILISVGAACHMCHVAHSEFEDPVAAMNSGFVCKVDREERPIDIVYNNAT
VALTGHGWPMTCLPTPDGRPPFCGTYVKAFLDILSVATWRRRGEVEEASRI
AGELRMTSGLPAGCADLATALCDENVAIILNDRDVYHGGFSGAPFPSSALIBALIR
STERTGSAVTLQAVATGAMARGGITDQGGGFAYISVDMAMVPHFERKMYDNLIL
LRLVAMWARTCDPLAARVYATARFLADLDGMFTSLDADADRGSTYVWTPA
QUTLEVJGPDGMAAVPVGTEGTGEHGSVLRUPADPPTPRFDRVRAALLAARG
RIOPGRDQVYTSNMGAIITAFTEASVALDDGDLADARCARSVLIDLHVQGRBRA
SLGGVVGDSAAIILBDHMTATGLIALYOLSAEDSWLTATDILDTLALHAFDRPGR
WPTTADDAERLRLRPADPLDGAFTPSGASVTEALITLALHVGGERERYLRATTDALR
AAVALIERAPRTAGHMLAVAEANVRGFLDIAVACDPARALADARLAPGAGIIVGG
PRDSSELLAGRDVAGADAAYVCRGVCDLPVTRPELAAALGAPSP"
    complement (11151..11471)
    /locus_tag="MAP2707c"
    complement (11151..11471)
    /locus_tag="MAP2707c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA050504.1"
    /db_xref="GI:41397164"
    /translation="NMHLIVLSGAVADGAPHHTGPDPRGRASPVGLLIVYLLISTLFL
IASMRQLAKVESPDPQHPERPDQABESTDAVDADADADADGPDHANGSPRPBGSD
DEPG"
    complement (11468..11989)
    /locus_tag="MAP2708c"
    /db_xref="GI:41397164"
    /translation="NMHLIVLSGAVADGAPHHTGPDPRGRASPVGLLIVYLLISTLFL
IASMRQLAKVESPDPQHPERPDQABESTDAVDADADADADGPDHANGSPRPBGSD
DEPG"
    complement (11468..11989)
    /locus_tag="MAP2708c"
```

```
Query Match      14.6%; Score 87; DB 1; Length 301332;
Best Local Similarity 63.8%; Pred. No. 2.8e-05;
Matches 132; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

gene
    complement (11468..11989)
    /locus_tag="MAP2708c"

20  CGGCGCCGAGCGCGCAGCACTGCTGCACATCTCGCCCGGTAGCAGGTGCGTTGAGGC 79
Db  245970  CGCGCGCGCGCGCGCGCGCACCTGCTGCAGATCTCGGCCCTTCACAGGTTGCGGCTCAGCC 246029
Qy  80  CGTGCGCCGCCGACGCGCAGAACTGGCAGCGGAGCGCACTCCGCTTGAAGACGAGACGC 139
Db  246030  CGCCCTGCGCGCGCGCGCGCAGAAAGGCGAGCCATGCGCGAGCGGCGCTGGAAGATGC 246089
Qy  140  AGCGCGGTGAGCGCTGCTGTCGACCGATCCGACCGCTTCGATGAATGCGCTCGACCA 199
Db  246090  ACACCGTGTGCGGTGCGGTGAGCGCATGACGACGATCCAGGTGACCCATCCAGGG 246149
Qy  200  GCTCGAAGCAAGAACTTGTGCTGTGCG 226
Db  246150  CCGCGCAGCAAGCTTTCGGGTCTGCG 246176
```

```
RESULT 3
BA000045_22
WPCOMMIT
Sequence split into 47 fragments LOCUS BA000045 Accession BA000045
Fragment Name      Begin      End
BA000045_00        1          110000
BA000045_01        100001     210000
BA000045_02        200001     310000
BA000045_03        300001     410000
BA000045_04        400001     510000
BA000045_05        500001     610000
BA000045_06        600001     710000
BA000045_07        700001     810000
BA000045_08        800001     910000
BA000045_09        900001    1010000
BA000045_10       1000001    1110000
BA000045_11       1100001    1210000
BA000045_12       1200001    1310000
BA000045_13       1300001    1410000
BA000045_14       1400001    1510000
BA000045_15       1500001    1610000
BA000045_16       1600001    1710000
BA000045_17       1700001    1810000
BA000045_18       1800001    1910000
BA000045_19       1900001    2010000
BA000045_20       2000001    2110000
BA000045_21       2100001    2210000
BA000045_22       2200001    2310000
BA000045_23       2300001    2410000
BA000045_24       2400001    2510000
BA000045_25       2500001    2610000
BA000045_26       2600001    2710000
BA000045_27       2700001    2810000
BA000045_28       2800001    2910000
BA000045_29       2900001    3010000
BA000045_30       3000001    3110000
BA000045_31       3100001    3210000
BA000045_32       3200001    3310000
BA000045_33       3300001    3410000
BA000045_34       3400001    3510000
BA000045_35       3500001    3610000
BA000045_36       3600001    3710000
BA000045_37       3700001    3810000
BA000045_38       3800001    3910000
BA000045_39       3900001    4010000
```

BA000045_40 4000001 4110000
BA000045_41 4100001 4210000
BA000045_42 4200001 4310000
BA000045_43 4300001 4410000
BA000045_44 4400001 4510000
BA000045_45 4500001 4610000
BA000045_46 4600001 4659019
Continuation (23 of 47) of BA000045 from base 2200001 (BA000045 Gloeobacter violaceus PC

Query Match 14.1%; Score 84.4; DB 1; Length 110000;
Best Local Similarity 58.3%; Pred. No. 9e-05;
Matches 148; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 3 CACAGAGCTGTTGACGCGGCGCCGAGCGCAGACCTGTGACGATCTGCGCCGCTAG 62
D 31182 CACCCGCGCGCGCAAAACCGGATTGCGCGGTGGCGCAAGAGCGGCGCATGGGACAACC 31241
QY 63 CAGGTGTGCTTGAAGCGCGTGCAGCGCGCAAGAACTGGACCGCGAAAGCGCACCC 122
D 31242 CAGGTGCGGGCGCAAAACCGGATTGCGCGGTGGCGCAAGAGCGGCGCATGGGACAACC 31301
QY 123 CGCTTGAGACGACGACGAGGCGGTGTACCCCTGCTGCGACGATTCGCAACCGCTGAT 182
D 31302 CACCTGCGACGAAACGACGACCGTCAAGCCGCTGCGCGCGCAATACCGAGTCTTGAC 31361
QY 183 GAATTTGCGCTGACGACGAGTCAAGAACTTTGTGTGCTGCTCCCTGCTGCGACT 242
D 31362 GATTTCGCGTGGCGCCCGCGCAAGAGTACTTATCTGCTGCGCGCGCATGTTG 31421
QY 243 GCGCTCGCGCGAGGG 256
D 31422 GCGACACGACCTGCG 31435

RESULT 4
SC039124/c 308050 bp DNA linear BCT 16-APR-2005
LOCUS SC039124
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 21/29.
ACCESSION AL039124 AL039129 AL031031 AL031035 AL031124 AL031225 AL031231
AL031260 AL034447 AL035559 AL035569 AL355913 AL645882
VERSION AL39124.1 GI:2441386
KEYWORDS
SOURCE
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kleiser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kleiser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorek,A., Woodward,J., Barrrell,B.G.,
Parkhill,J. and Hopwood,D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
PUBMED 1200953
REFERENCE 2 (bases 1 to 308050)
AUTHORS Bentley,S.D.
JOURNAL Direct Submission

COMMENT Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridgeshire CB10 1SA E-mail: sds@sanger.ac.uk
On or before Oct 26, 2002 this sequence version replaced
gi:20500664, gi:20500751, gi:20520752, gi:20520665, gi:20520811,
gi:20520756, gi:20520845, gi:20520760, gi:20520817, gi:20520818,
gi:20520782.

FEATURES
source 1. 308050
/organism="Streptomyces coelicolor A3(2)"

gene
CDS

/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
250..2490
/gene="SC05495"
/note="synonym: SC8D9.07"
250..2490
/gene="SC05495"

/note="SC8D9.07, possible membrane associated
phosphodiesterase, len: 746 aa; similar to many eg.
TR-O87376 (EMBL:AF052518) c-di-GMP phosphodiesterase A
from Acetobacter xylinus cdg2 operon (752 aa) fasta
score: 959, z-score: 1059.2, E(): 0, (31.7% identity
in 603 aa overlap). Appears to have two distinct regions
with the first 300aa containing several potential membrane
spanning hydrophobic regions and the remainder of the
protein showing similarity to the phosphodiesterase
mentioned above. Also similar to a nearby gene on the same
cosmid (SC8D9.22) fasta scores: opt: 736, z-score: 625.6,
E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains
TVA leu codon, possible target for bldA regulation. Also
contains Pfam match to entry PF00990 DUF9, Domain of
unknown function, score 166.70, E-value 3.9e-46 and Pfam
match to entry PF00563 DUF2, Domain of unknown function 2,
score 291.30, E-value 1.2e-83. Contains possible
hydrophobic membrane spanning regions"

/codon_start=1
/transl_table=11
/product="putative phosphodiesterase"

/protein_id="CA837572.1"

/db_xref="GI:4467246"

/db_xref="GOA:Q92584"

/db_xref="InterPro:IPR001633"

/db_xref="UniProt/TREMBL:Q92584"

/transl_table=11
/product="putative phosphodiesterase"
/protein_id="CA837572.1"
/db_xref="GI:4467246"
/db_xref="GOA:Q92584"
/db_xref="InterPro:IPR001633"
/db_xref="UniProt/TREMBL:Q92584"
/transl_table=11
/product="putative phosphodiesterase"
/protein_id="CA837572.1"
/db_xref="GI:4467246"
/db_xref="GOA:Q92584"
/db_xref="InterPro:IPR001633"
/db_xref="UniProt/TREMBL:Q92584"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

/db_xref="UniProt/TREMBL:Q9Z583"
/translation="MTVATYLAIVLNLFCIFLGYRFLQFPASATGYGYPADAGDA
AAYLVKGVSDGTGGVGLALVLAAPAGARPEAMFGLCVALVPLADTLVLNRGGTKAVA
EGIHFAIYVLLSAGLLFAV"
3177..3671
/gene="SC05497"
/note="synonym: SC8D9.09"
3177..3671
/gene="SC05497"
/note="SC8D9.09, unknown, len: 164aa; similar to many of
undefined function eg. TR:006336 (EMBL:Z95390)
hypothetical protein from Mycobacterium tuberculosis (177
aa) fasta scores; opt: 171, z-score: 213.4, E(): 0.00014,
(34.9% identity in 109 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB37573.1"
/db_xref="GI:4467248"
/db_xref="GOA:Q9Z582"
/db_xref="InterPro:IPR003313"
/db_xref="InterPro:IPR007113"
/db_xref="UniProt/TREMBL:Q9Z582"
/translation="MSLVKPTDSEVVVREAEAVNGRAPTVRLADSSSTGALST
QRYTLTAGDAGAKPHMDNSAEMLFDGADYLSGDEVLTAGGDLIVPPKPHAF
AAVGGADADLLIYLAPEGVERFEYFRHLQRIALGATPESLLEVQELVDNHLSSAAMD
ARRG"
3816..4112
/gene="SC05498"
/note="synonyms: gacC, SC8D9.10"
3816..4112
/gene="SC05498"
/note="SC8D9.10, probable Glu-tRNA-Gln amidotransferase
subunit C, gacC, len: 98aa; similar to many eg. TR:006492
(EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit C,
gacC, from Bacillus subtilis (101 aa) fasta scores; opt:
232, z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95
aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA-Gln amidotransferase subunit C"
/protein_id="CAB37574.1"
/db_xref="GI:4467249"
/db_xref="GOA:Q9Z581"
/db_xref="InterPro:IPR003837"
/db_xref="InterPro:IPR004415"
/db_xref="UniProt/Swiss-Prot:Q9Z581"
/translation="MPQITREYVAHLARLRLKCBELHPAGQLDDIIGAVARVSE
VADQVPPTSHPLPLTNVMPDEVRPSLTPBQALSGAPAQEQRFKVPQILGSE"
4116..5609
/gene="SC05499"
/note="synonyms: gacA, SC8D9.11"
4116..5609
/gene="SC05499"
/note="SC8D9.11, probable Glu-tRNA-Gln amidotransferase
subunit A, gacA, len: 497aa; similar to many eg. TR:006491
(EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit A,
gacA, from Bacillus subtilis (486 aa) fasta scores; opt:
1443, z-score: 1512.5, E(): 0, (52.0% identity in 465 aa
overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA-Gln amidotransferase subunit"
/protein_id="CAB37575.1"
/db_xref="GI:4467250"
/db_xref="GOA:Q9Z580"
/db_xref="InterPro:IPR000120"
/db_xref="InterPro:IPR004412"
/db_xref="UniProt/Swiss-Prot:Q9Z580"
/translation="MSDIKTLTAIAKIASGETLAVOVTEAHARIEAVDEKTHAF
LHVDRGALAQAQAVDEKRGSKGLGVLALADITPTTBEIPPTVSKLIEGIP
PDYATVTKRLKADVVILGKTNDBFAMGSSSTENSAVGTGNFMDLTKIPGSGGSS
AALAFQAPLAIGTDTGSGIRQPAVYTGVPKTYGVSRYGMVAFSSSLDGGGRCA

RVTLDAALLHEVIAGHDPMDDSTSIDAPVPAVVEAARNGSVGNRGVVKOPRGEYOA
GVVORDBSEIETLSKSGAIEVELDCSPDLASAYLIASBSSNARGLARAGAR
VGDGCHNSAEETSLIRBAGREKVRIRMLGTYALSSGYDAYGAQRRTIKRBS
FERAPROVDIVPTPTTTFALGERADDPMANTLADLCTIPNLGAAMSLPCGLA
PEDNLPVGLQIILAPAMKDRLYGVGAVEAFAVEKWHPLIBRAPSL"
5606..5845
/gene="SC05500"
/note="synonym: SC8D9.12"
5606..5845
/gene="SC05500"
/note="SC8D9.12, putative membrane protein, len: 79 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB37576.1"
/db_xref="GI:4467251"

Query Match 13.9%; Score 83.2; DB 1; Length 308050;
Best Local Similarity 62.5%; Pred. No. 0.00012;
Matches 130; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 14 TGAAGCCGCGGCCCGAGCGCCAGACACTGTCTGACGATCTGCCCGGTAGCAGTGTGCT 73
Db 162508 TGAAGCGCCGCGCATGCGCGTCCACGATCTGGTGACGATCTCGCGGTGACAGGTTGCGGT 162449
QY 74 TGAAGCCGCGCGGCCCGAGCGCCAGACACTGTCTGACGCGGAAGCGGCACCCCGCTGAGACG 133
Db 162448 CGAGGCGCGCGCTGCGCGGTGCGCCAGAACCGGAGTTCACTCCGCGCGCGCTGAGACG 162389
QY 134 AGACGAGCGCGCGGTGAGCGGTGCGTGGAGACGATCCGCGCGGTGATGAATGCGCGT 193
Db 162388 TGAATGACATGATGACACCCCGGTCCGGTAGGACATGAGACCGACCTGACGAGCGTCCGT 162329
QY 194 CGACGAGCTCGAACAGAACTTGTGCT 221
Db 162328 CGAACGCTTCCACAGCGTCTTCCGCGT 162301

RESULT 5
AL646063/LOCUS 207050 bp DNA linear BCT 16-APR-2005
DEFINITION Ralstonia solanacearum GM11000 chromosome, complete sequence;
segment 7/19
ACCESSION AL646063 AL646052
VERSION AL646063.1 GI:17428161
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
1
REFERENCE
AUTHORS Slatonubart,M., Genin,S., Artiguenave,F., Gouzy,J., Mangent,S.,
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunac,S., Demange,N.,
Caspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siglier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Nature 415 (6871), 497-502 (2002)
11823852
PUBMED
TITLE Nature 415 (6871), 497-502 (2002)
JOURNAL
REFERENCE
PUBMED 2 (baes 1 to 207050)
Boucher,C.A.
AUTHORS Direct Substition
TITLE Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CERH, 27 rue Juliette Dodu, 75010 Paris, France, IGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire

COMMENT INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Bouchere@coulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES
Source
1. 207050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
complement(332..1561)
/gene="RSc1148"
/note="synonym: RS04604"
complement(332..1561)
/gene="RSc1148"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD14850.1"
/db_xref="GI:17428162"
/db_xref="InterPro:IPR003347"
/db_xref="InterPro:IPR007113"
/db_xref="UniProt/TREMBL:Q8Y096"
/translation="MNELPLVAAGVPGIAGAAGPAQLGSLSPDFMRTWKKPL
LIRQALSPEDKRALHPLSPALIKLARREVSRLLAQTRGRTFNGNPNRPLPS
RKARWTLTVGVNLVPAVDALMORFRFIPDARDVMI SPATGSGVGHPSYDV
PLIAGHGRRRIRISONDITLVPDLPLILNFPBESFTLEPGMLTLPHYADGV
ARQDMTVSGFRSPSYRELAGHFGFISQLENDPNPDREGYTPDEQCPVRRPLPT
AMVRALQKLNALNTPRLVGSFFGAVITPEPDHVEFTQPLSLARTAAKRGIV
LDARTQALYDAQRFVINGDTPEPSDTLLAMLSLSDKAGAAVAADAADLPDLMDTL
HAYEEGWLQLAGPGAH"
1573..2097
/gene="alyD"
/note="synonym: RS04602, RSc1149"
1573..2097
/gene="alyD"
/BC_number="5.2.1.8"
/function="cell processes; chaperoning"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=1
/product="PROBABLE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS
ISOMERASE (ROTAMASE) PROTEIN"
/protein_id="CAD14851.1"
/db_xref="GI:17428163"
/db_xref="GOA:Q8Y095"
/db_xref="InterPro:IPR001179"
/db_xref="UniProt/TREMBL:Q8Y095"
/translation="MKIANTVSAVAYKLSDAQGVIESDEPMVYLHGVDGTPPKI
BEALDGHGEGFETKLEPDAFGEYDQELIKERDRPPEBLEVGMOFEGVGEQDD
BDALITVTVDAVEDKVVLDGNHPLAGMALRLPRLARTVDRVAAEEIEGHVHVGEGLE
VADDEDEPPRTIH"
complement(2152..3216)
/gene="RSc1150"
/note="synonym: RS04601"
complement(2152..3216)
/gene="RSc1150"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"

/codon_start=1
/evidence-not_experimental
/transl_table=1
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD14852.1"
/db_xref="GI:17428164"
/db_xref="UniProt/TREMBL:Q8Y094"
/translation="MARPGAFAFRMRIGIALMLATAAPAPKGHATAPAAPACE
TFVAVGVNPPQRPEDVAPARALDAIDHEPAFVHLGNLGRDESCTDNLIERHDL
DSITPLIYVIGDHNDGCGPRTAGRPDLRLRLRLDLFPDNLGQRTVMRQS
DQAKFRSPRNARMTGGVLFYTLNVGDNNTKTAGRGGEYEDLEAROMYARAF
ALNRKLPQPMVTIMQADPLPDEGMRERRPNLLDGLRRRPHDGYALRQGLARL
EYDGAVLILHGASQSFILDRPLDQESGRPDHPHVARFQSPFTLDQMLEVSPGTAP
LFHIRARILASDIAPSTPQQ"
complement(3216..5864)
/gene="mts"
/note="synonym: RS04600, RSc1151"
complement(3216..5864)
/gene="mts"
/function="macromolecule metabolism; macromolecule
synthesis, modification; dna - replication, repair,
restr./modif."
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=1
/product="PROBABLE DNA MISMATCH REPAIR PROTEIN"
/protein_id="CAD14853.1"
/db_xref="GI:17428165"
/db_xref="GOA:Q8Y093"
/db_xref="InterPro:IPR000432"
/db_xref="InterPro:IPR005748"
/db_xref="InterPro:IPR007695"
/db_xref="InterPro:IPR007696"
/db_xref="InterPro:IPR007860"
/db_xref="InterPro:IPR007861"
/db_xref="UniProt/Swiss-Prot:Q8Y093"
/translation="WASHAPALFIDKYGFPDRTPMQOYLKSNHPTTVPYRM
GDFYLPFEDAEEKARLLDITLTSRGTTNGQPIHMAQVPFAVQYLAKLVLESVA
ICQIGDPATKGVKRVKVVVPTGLTALALSDVNNHLLAIQIPGRGAAPLV
GLANLVVGGELRLMECSAQDLRELRIRAEVLAADATLNAVAFVARRLPDMHF
DVEAGARRIREQIVASVAGCETLTLAALAAAGALNVAATGQSGIRVYIGLVSH
ESRIGIDTATRRRLLETTRTQGESPTLSELDTCAITMSGRLLRHTLHPLDRAY
PQARQAIIVLADGMSLKTGLTLDVERITRRLALMSARPDDLSIRTLARPE
VRDQLPQSDAQGLDILAAALAIPEDAHALQRAVMAEPMAVRDGGVIAQVADLD
ELRDISRCQFLVDLBARERERTGIANLREYVRVHGFIETNGQAAKVPDYRRR
QTLKNAERYITPELKAFFEDKALSAQDRLSREKLYEDLQKLPHLFVLRIAAALA
QADVIATLAAAHALMSRPALTDAPGIELRABHPVYQOVQEVANDCVLQSTRKL
LITGPNMGKSTFMRCATVYVLAAYGAPVPAATGPIDRIETRIGADDLAAGR
STFVETMEAAAILHRAIPNSLVIMDEIGRTSTFDGIALAMALARHLSHNRSHITLF
ATHYFELLQLPQEPRAQAAVHLISAVERHDGIVFLHAAVQEGPASQSYGIVQVLGVQ
PVIRARAKRLAMEQHSADTGATPOLDLFALPSDDDAEAAAPAPSALEALDG
IDPSPMPRDAIDALYRLKALDSASA"
complement(5960..6853)
/gene="RSc1152"
/note="synonym: RS04599"
complement(5960..6853)
/gene="RSc1152"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=1
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14854.1"

AE000516_08 800001 910000
AE000516_09 900001 1010000
AE000516_10 1000001 1110000
AE000516_11 1100001 1210000
AE000516_12 1200001 1310000
AE000516_13 1300001 1410000
AE000516_14 1400001 1510000
AE000516_15 1500001 1610000
AE000516_16 1600001 1710000
AE000516_17 1700001 1810000
AE000516_18 1800001 1910000
AE000516_19 1900001 2010000
AE000516_20 2000001 2110000
AE000516_21 2100001 2210000
AE000516_22 2200001 2310000
AE000516_23 2300001 2410000
AE000516_24 2400001 2510000
AE000516_25 2500001 2610000
AE000516_26 2600001 2710000
AE000516_27 2700001 2810000
AE000516_28 2800001 2910000
AE000516_29 2900001 3010000
AE000516_30 3000001 3110000
AE000516_31 3100001 3210000
AE000516_32 3200001 3310000
AE000516_33 3300001 3410000
AE000516_34 3400001 3510000
AE000516_35 3500001 3610000
AE000516_36 3600001 3710000
AE000516_37 3700001 3810000
AE000516_38 3800001 3910000
AE000516_39 3900001 4010000
AE000516_40 4000001 4110000
AE000516_41 4100001 4210000
AE000516_42 4200001 4310000
AE000516_43 4300001 4403837

Continuation (32 of 44) of AE000516 from base 3100001 (AE000516 Mycobacterium tuberculosis)

Query Match 13.0%; Score 77.8; DB 1; Length 110000;
Best Local Similarity 61.7%; Pred. No. 0.0011;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 26 CGAGCGCAGCACTGTGACGATCTGCCCGGTAGCAGGTGTGCTTGAGGCCGTGC 85
Db 8434 CGCCGCGACGACCTGCTCGAGGATCTCCGCTGATAGGTGGAATCATCCGCCCT 84423
Qy 86 GAGCCGACGCGCAGAACTGGACGCGGAGGCGCACCCGCTTGAGACGACGCGCG 145
Db 84424 GAGCCGATGGACAGAAAGCGGCAAGCCATGCGCGAGCCGCTGCGAGAAATGCGACCG 84483
Qy 146 TGTAGCCGTGTGGGAGGATTCGCAACCGTCTGATGAATTCGCCGTGACGACGTGGA 205
Db 84484 TGTGTGCGCGCGGATAGGCGATTCGACACCGATTCGACATGTGATCCGTGACGCGCCGCC 84543
Qy 206 ACAGGAATTGTGCTGTGC 226
Db 84544 ACAAGTCTTTCAGTCTGCC 84564

RESULT 8
AE017333_17/c
WPCOMMENT
Sequence split into 43 fragments LOCUS AE017333 Accession AE017333

Fragment Name	Begin	End
AE017333_00	1	110000
AE017333_01	100001	210000
AE017333_02	200001	310000
AE017333_03	300001	410000
AE017333_04	400001	510000
AE017333_05	500001	610000
AE017333_06	600001	710000
AE017333_07	700001	810000
AE017333_08	800001	910000

AE017333_09 900001 1010000
AE017333_10 1000001 1110000
AE017333_11 1100001 1210000
AE017333_12 1200001 1310000
AE017333_13 1300001 1410000
AE017333_14 1400001 1510000
AE017333_15 1500001 1610000
AE017333_16 1600001 1710000
AE017333_17 1700001 1810000
AE017333_18 1800001 1910000
AE017333_19 1900001 2010000
AE017333_20 2000001 2110000
AE017333_21 2100001 2210000
AE017333_22 2200001 2310000
AE017333_23 2300001 2410000
AE017333_24 2400001 2510000
AE017333_25 2500001 2610000
AE017333_26 2600001 2710000
AE017333_27 2700001 2810000
AE017333_28 2800001 2910000
AE017333_29 2900001 3010000
AE017333_30 3000001 3110000
AE017333_31 3100001 3210000
AE017333_32 3200001 3310000
AE017333_33 3300001 3410000
AE017333_34 3400001 3510000
AE017333_35 3500001 3610000
AE017333_36 3600001 3710000
AE017333_37 3700001 3810000
AE017333_38 3800001 3910000
AE017333_39 3900001 4010000
AE017333_40 4000001 4110000
AE017333_41 4100001 4210000
AE017333_42 4200001 4222645

Continuation (18 of 43) of AE017333 from base 1700001 (AE017333 Bacillus licheniformis)

Query Match 13.0%; Score 77.8; DB 1; Length 110000;
Best Local Similarity 63.8%; Pred. No. 0.0011;
Matches 116; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 37 ACCTGTGACGATCTGCCCGGTAGCAGGTGTGCTTGAGAGCCGTGCGCCGAGCG 96
Db 55396 ACCTGTGACGATCTGCCCGGTAGCAGGTGTGCTTGAGAGCCGAGGTGATGCG 55337
Qy 97 CAGAACTGACGCGGAGGCGCACCCGCTTGAGACGACGCGGTGTGACCGTGC 156
Db 55336 CAGAACTGACGCGGATCCGCGACCCGCTGTGTCTTACCAAAAGAAATGCCATAT 55277
Qy 157 TGGCAGGATTCGCAACCGTCTGATGAATTCGCCGTGACGACCTGGAACGAACTTT 216
Db 55276 TCGTGGCGCATTAATAGGTCTGATCGTGTAGCCGTCATGAAGCTGAAATAAAACCTTC 55217
Qy 217 GTGCT 221
Db 55216 ATTGT 55212

RESULT 9
CP000002_17/c
WPCOMMENT
Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000

CP000002.11 1100001 1210000
 CP000002.12 1200001 1310000
 CP000002.13 1300001 1410000
 CP000002.14 1400001 1510000
 CP000002.15 1500001 1610000
 CP000002.16 1600001 1710000
 CP000002.17 1700001 1810000
 CP000002.18 1800001 1910000
 CP000002.19 1900001 2010000
 CP000002.20 2000001 2110000
 CP000002.21 2100001 2210000
 CP000002.22 2200001 2310000
 CP000002.23 2300001 2410000
 CP000002.24 2400001 2510000
 CP000002.25 2500001 2610000
 CP000002.26 2600001 2710000
 CP000002.27 2700001 2810000
 CP000002.28 2800001 2910000
 CP000002.29 2900001 3010000
 CP000002.30 3000001 3110000
 CP000002.31 3100001 3210000
 CP000002.32 3200001 3310000
 CP000002.33 3300001 3410000
 CP000002.34 3400001 3510000
 CP000002.35 3500001 3610000
 CP000002.36 3600001 3710000
 CP000002.37 3700001 3810000
 CP000002.38 3800001 3910000
 CP000002.39 3900001 4010000
 CP000002.40 4000001 4110000
 CP000002.41 4100001 4210000
 CP000002.42 4200001 4222334

Continuation (16 of 43) of CP000002 from base 1700001 (CP000002 Bacillus licheniformis A

Query Match 13.0%; Score 77.8; DB 1; Length 110000;
 Best Local Similarity 63.8%; Pred. No. 0.0011;
 Matches 118; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 37 ACCTGTCGACGATTCGCCCCGTTAGAGGTGCTGCTTGAAGCCGCTGCGCCCGACGCG 96
 Db 54550 ACCTGTGCGACGATCTCTCCGGCTTCAGAGTCCGCTTTAAGCCGCGAGGCTGATGCG 54491
 QY 97 CAGAACTGCGACCGAAGGCGCAACCCGCTGAGAGAGAGAGAGGCGGCTGAGCGTCG 156
 Db 54490 CAGAACTGACGCGGATCCGCGAGCCGACCTGTGTCTTCAACAACGAAATTCGCAATAT 54431
 QY 157 TGGCGACGATTCGCGACCGCTCTCGATGAATTCGCGCTGACGACCTCGAAGCAACTTT 216
 Db 54430 TCGTGCCTGATTAATACGCTCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 54371
 QY 217 GTGCT 221
 Db 54370 ATTGT 54366

RESULT 10
 BX248344 318050 bp DNA linear BCT 17-Apr-2005
 LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 11/14.
 ACCESSION BX248344 BX248333
 VERSION BX248344.1 GI:31619628
 KEYWORDS complete genome.
 SOURCE Mycobacterium bovis AF2122/97
 ORGANISM Mycobacterium bovis AF2122/97
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

REFERENCE
 AUTHORS Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
 Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsempé,C., Simon,S.,
 Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
 Wheeler,P.R., Parhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and

TITLE Hewinson,G.
 JOURNAL The complete genome sequence of Mycobacterium bovis
 REFERENCE Unpublished
 2 (bases 1 to 318050)
 AUTHORS Garnier,T.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2003) Garnier T., Unité de Génétique Moléculaire
 Bactérienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
 Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
 Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
 Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
 Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
 Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28
 rue du Docteur Roux, 75724 Paris Cedex 15, France
 location/Qualifiers
 1. 318050
 /organism="Mycobacterium bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 /complement(99. 875)
 /locus_tag="Mb2882c"
 /complement(99. 875)
 /locus_tag="Mb2882c"
 /BC_number="1.1.1.1"
 /note="Mb2882c, -, len: 258 aa. Equivalent to Rv2857c,
 len: 258 aa, from Mycobacterium tuberculosis strain H37Rv,
 (100.0% identity in 258 aa overlap). Probable short-chain
 dehydrogenase/reductase (EC 1.1.1.1), highly similar to
 various dehydrogenases e.g. O88068|SC15.33c PROBABLE
 DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
 (260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
 identity in 253 aa overlap); Q91376|PA1649 from
 Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
 (253 aa), FASTA scores: opt: 569, E(): 2.1e-28, (39.2%
 identity in 255 aa overlap); Q9EX74|MLHA SDR-LIKE ENZYME
 from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
 567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);
 etc. Also similar to many Mycobacterium tuberculosis
 dehydrogenases e.g. FABG3|Rv2002|MT0581|MT039.16c
 PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%
 identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
 DEHYDROGENASES/REDUCTASES (SDR) FAMILY."
 /codon_start=1
 /transl_table=11
 /product="PROBABLE SHORT-CHAIN TYPE
 DEHYDROGENASE/REDUCTASE"
 /protein_id="CAD96569.1"
 /db_xref="GI:31619629"
 /db_xref="GOA:Q7TXP6"
 /db_xref="InterPro:IPR002198"
 /db_xref="InterPro:IPR002347"
 /db_xref="UniProt/TREMBL:Q7TXP6"
 /translation="MMDLSQRLAGRAVITGGSGIGLAGRRMRAGATIVGDVDV
 EAGGAADSLGLFPTDVCDDDAVNGLDGAGATGRIDIFNNNGISPPBNLLN
 TEIAAWORVDVNLKSVYLCRAALRHMTLAGSIVNPAFAVNGSATSGSYAS
 KCGVLAWSRELGVOPAROGIRVNALCPGVNTPDLDELPAKRNPERAARAHVHPLERF
 APPDRIAAVAFIASDASFTASTFLVDVGISAAVTVPL"
 /complement(872. 2239)
 /gene="aldC"
 /locus_tag="Mb2883c"
 /complement(872. 2239)
 /gene="aldC"
 /locus_tag="Mb2883c"
 /complement(872. 2239)
 /BC_number="1.2.1.3"
 /note="Mb2883c, aldC, len: 455 aa. Equivalent to Rv2858c,
 len: 455 aa, from Mycobacterium tuberculosis strain H37Rv,
 (100.0% identity in 455 aa overlap). Probable aldC,
 aldehyde dehydrogenase (EC 1.2.1.3), similar to many e.g.
 O88069|SC13.34c PUTATIVE ALDEHYDE DEHYDROGENASE from
 Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,

E(): 6.4e-109, (64.5% identity in 448 aa overlap);
O9FAB1|ALPH|BT-ALDH ALDEHYDE DEHYDROGENASE from *Bacillus*
thermoautotrophicus (497 aa), FASTA scores: opt: 1157, E():
2.1e-64, (44.3% identity in 458 aa overlap); O33455|CMC
P-CUMIC ALDEHYDE DEHYDROGENASE from *Pseudomonas putida*
(494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15%
identity in 452 aa overlap);
P40047|DHAS YEAST|ALDH5|ALD5|YER073M ALDEHYDE
DEHYDROGENASE from *Saccharomyces cerevisiae* (Baker's
yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60,
(38.55% identity in 459 aa overlap);
P80668|FEAB_ECOLI|PADA|MA08|B1385 PHENYLACETALDEHYDE
DEHYDROGENASE (EC 1.2.1.39) from *Escherichia coli* strain
K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2%
identity in 462 aa overlap); etc. Also similar to many M.
tuberculosis dehydrogenases e.g. P71823|Rv0768|MT0363.13
(489 aa), FASTA score: (38.1% identity in 467 aa overlap).
Contains P500687 Aldehyde dehydrogenases glutamic acid
active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES
FAMILY."

/codon_start=1
/transl_table=1
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"
/protein_id="CAD96570.1"
/db_xref="GI:31619630"
/db_xref="GOA:O7TXP5"
/db_xref="InterPro:IPR020866"
/db_xref="UniProt/TrEMBL:O7TXP5"
/translation="MSTTQLINPAETVEVLASVDHTANAVDAVQRAAQRNRARIA
PAQRAAGIRAPAAVAQALDELIALLEVANSCHPIVSAMEAGHRAVDIAFAAPSEL
SGROI PVAGVDVTFENPMGVGVIITPNNPMTVASIALPALAGNAVIAAPDEL
LTMRIEGLAVBAGLDLLOVLPGKGVVGERPVTHDIKIVFTSGTEKGVKVMG
AAAYKVITLIEGKSKANIVFHDCLDRAATAPAGVDNAGDCCASRLIVASVY
DRFELIEPAVHSTIVGDPGSRATKEMGLVSRANDKAYGTPDAPAPRGTAPRGR
GFMPFPVTLTPKRDRTVDEIFGPVVVLIFFDEADAIISLANDIAYGLSSIVTDL
SRALRVAAVSQGNLSVNSHSVRFNTFPFGFKQSGVGRLEIGDAPLQFTKXVFLA
VGEEM"

complement (2236, 3162)
/locus_tag="MD2884c"
complement (2236, 3162)
/locus_tag="MD2884c"
/EC_number="2.-.-.-"
/EC_number="6.3.5.-"
/note="MD2884c", len: 308 aa. Equivalent to Rv2859C.
len: 308 aa, from *Mycobacterium tuberculosis* strain H37Rv,
(100.0% identity in 308 aa overlap). Possible
amidotransferase (EC 6.3.5.- or 2.-.-.-), equivalent (but
longer 58 aa) to Q9CBU9|ML1573 POSSIBLE AMIDOTRANSFERASE
from *Mycobacterium leprae* (249 aa), FASTA scores: opt:
1226, E(): 3e-64, (71.55% identity in 239 aa overlap).
Also similar to other amidotransferases and hypothetical
proteins, but shorter in N-terminus e.g. O88072|SC135.37
HYPOTHETICAL 25.3 KDA PROTEIN from *Streptomyces coelicolor*
(242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65%
identity in 235 aa overlap); AAK79730|Q97186|CAC1764
PREDICTED GLUTAMINE AMIDOTRANSFERASE from *Clostridium*
acetobutylicum (241 aa), FASTA scores: opt: 458, E():
1.6e-19, (32.95% identity in 246 aa overlap);
AAK75201|Q97QV9|SP1089 GLUTAMINE AMIDOTRANSFERASE CLASS I
from *Streptococcus pneumoniae* (229 aa), FASTA scores: opt:
431, E(): 5.6e-18, (34.75% identity in 236 aa overlap);
etc. Contains three 17 aa repeats at the N-terminus very
similar to those in other *Mycobacterium tuberculosis*
proteins e.g. Q10699|YJ30_MCTU|Rv2090|MT2151|MT049.30
PUTATIVE 5'-3' EXONUCLEASE Rv2090 (EC 3.1.11.-)."
/codon_start=1
/transl_table=1
/product="POSSIBLE AMIDOTRANSFERASE"
/protein_id="CAD96571.1"
/db_xref="GI:31619631"
/db_xref="GOA:O7TXP4"
/db_xref="InterPro:IPR011697"

/db_xref="UniProt/TrEMBL:O7TXP4"
/translation="MDLSASRSRSGDPLRPASRLSPVSGDPLRPASRLSPVS
DGDPLRPASRLSPVGLTALVQVTRGWDIPAGLPAIDYFEGITMAG
VAVLPPQVPDPSVSGVDLSLALVITGGYDLDPLAAGQSPAPDPRPDAMFE
ALRGLORCMPTLGI CRGQVNVNVALGTHLHPILIGSGHAGNGVETRLPVHT
ASGTRIALELIGESADPCTHQAIDQVBSLVYSAVDVGVIALBELPGDTFLAVGM
HPRKSLDDLRLFPALVDASGVAGRSQAPR"
complement (3143, 4516)
/gene="glnA4"
/locus_tag="MD2885c"
complement (3143, 4516)
/gene="glnA4"
/locus_tag="MD2885c"
/EC_number="6.3.1.2"
/note="MD2885c, glnA4, len: 457 aa. Equivalent to Rv2860C,
len: 457 aa, from *Mycobacterium tuberculosis* strain H37Rv,
(100.0% identity in 457 aa overlap). Probable glnA4,
glutamate synthetase class II (EC 6.3.1.2), similar to
many glutamine synthetases e.g. O88070|SC135.35C from
Streptomyces coelicolor (462 aa), FASTA scores: opt: 1947,
E(): 8.2e-120, (64.15% identity in 452 aa overlap);
O98H15|ML3074 from *Rhizobium loti* (Mesorhizobium loti)
(465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7%
identity in 452 aa overlap); Q98EM0|ML4187 from *Rhizobium*
loti (Mesorhizobium loti) (456 aa), FASTA scores: opt:
698, E(): 4.6e-38, (33.5% identity in 454 aa overlap);
Q9CDL9|GLNA from *Lactococcus lactis* (subsp. *lactis*)
(Streptococcus lactis) (446 aa), FASTA scores: opt: 633,
E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.
Also similar to three other potential glutamine synthetases
in *Mycobacterium tuberculosis*:
Q10378|GLN2_MCTU|Rv2222c|MT2280|MT0190.33c|MT0427
. 0.3c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score:
. 31.1% identity in 453 aa overlap); Rv1878|glnA3 and
Rv2220|glnA1 BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY."
/codon_start=1
/transl_table=1
/product="PROBABLE GLUTAMINE SYNTHETASE GLNA4 (GLUTAMINE
SYNTHETASE)"

Query Match 13.0% Score 77.8; DB 1; Length 318050;
Best local similarity 61.7%; Pred. No. 0.00092;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY	26	CGAGCGCCGACGACCTGCTGCAGATCTGCCCGGATGACAGGTGCTTGAGCCGCTGC	85
DB	21725	CGCGGACGACGACCTGCTGCAGATCTGCCCGGATGACAGGTGCTTGAGCCGCTGC	21784
QY	86	GCGCCGACGCGACGATGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGG	145
DB	21785	GCGCAGTGGCACAGAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG	21844
QY	146	TGTAGCGCGTGTGGGACGATCCGACGATCCGATCCGATCCGATCCGATCCGATCCG	205
DB	21845	TGTAGCGCGCGGATGCGACGATCCGACGATCCGATCCGATCCGATCCGATCCGATCCG	21904
QY	206	ACAAGAACTTGTGCTGCTGCG	226
DB	21905	ACAAGCTTTTGAAGTCTGCG	21925

RESULT 11
BX842581 348676 bp DNA linear BCT 17-APR-2005
LOCUS BX842581
DEFINITION *Mycobacterium tuberculosis* H37Rv complete genome; segment 10/13.
ACCESSION BX842581 AL008883 AL008967 AL021070 AL021287 AL021309 AL123456
Z74024 Z74697 Z81331 Z83018 Z83857 Z83858 Z83866 Z95150 Z95207
VERSION BX842581.1 GI:41352756
KEYWORDS complete genome.
SOURCE *Mycobacterium tuberculosis* H37Rv
ORGANISM *Mycobacterium tuberculosis* H37Rv
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.


```

/protein_id="CAB03633.1"
/db_xref="GI:1648906"
/db_xref="GOL:Q50686"
/db_xref="InterPro:IPR002514"
/db_xref="InterPro:IPR009057"
/db_xref="UniProt/Swiss-Prot:Q50686"
/translation="MSGGSRARYPELIRERARVRAVRAIRGQDSEPMALISVAVILGV
GCAETRYKRVKRVQAVDAGARPGPTTBESHLELRRLRDNAELRRNALIKTSAPFAAE
LDRPAR"
complement(2845..2860)
/note="16 bp partial direct repeat, GTGCTCAGACCCAAA, of
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(2901..2936)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(2976..3011)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3049..3084)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3121..3156)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3193..3228)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3266..3301)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3338..3373)
/note="36 bp direct repeat, 32 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3399..3334)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3476..3511)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3548..3583)
/note="36 bp direct repeat, 36 out of 36 bp identical to
sequence GTGCTCAGACCCAAAACCCCGAGAGGGGAGCGAAAC"
complement(3624..3659)

Query Match      13.0%; Score 77.8; DB 1; Length 348676;
Beat Local Similarity 61.7%; Pred. No. 0.00091;
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

26 CGAGCGCCGACGACTCTGTCACAGATCTCGCCCGGTAGACAGTGTGCTTGAAGCCGTCCG 85
71173 CGCCGGCGACGACCTGCTCGAGATCTCGCCCGGTAGATAGTTCAGTCAATTCGCCCT 71232
86 GGCCCGACGGCGACGATGCGACGCGAAGGCGACCCCGCTGAGAGACGACGACGACG 145
71233 GGCCAGTGGCACAGAACGAGGAAACCATGCGACGCGCGCTGAGAGAAATCAACACG 71292
146 TGTAGCCCTCGTGGCGACGATTCGACACCGTCTCGATGAATTTGCGTTCGACCGAGCTCGA 205
71293 TGTTCGCGCGGAGATAGGCGATTCAGCACCGATTGAAACATGTATACGTTCGACGCGCGCG 71352
206 ACAGGAACCTTTGTGCTGCGC 226
71353 ACAACGTCTTCGAGTCTGGC 71373

RESULT 12
AR628287/c 1080 bp DNA linear PAT 14-FEB-2005
LOCUS AR628287
DEFINITION Sequence 9688 from patent US 6833447.
ACCESSION AR628287
VERSION AR628287.1 GI:59737208
KEYWORDS
SOURCE Unknown.
```

ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1080)
TITLE	Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.
JOURNAL	Myxococcus xanthus genome sequences and uses thereof
FEATURES	Patent: US 6833447-A 9688 21-DEC-2004,
source	Monanto Technology, LLC; St. Louis, MO
	Location/Qualifiers
	1..1080
	/organism="unknown"
	/mol_type="genomic DNA"
ORIGIN	
Query Match	13.0%; Score 77.6; DB 6; Length 1080;
Best Local Similarity	61.3%; Pred.No. 0.0026;
Matches 125; Conservative	0; Mismatches 79; Indels 0; Gaps 0;
Db	
Qy	34 AGCAGCTGCTGCACGATCTCGCCCGGTAGCAGGTGTGCTTGAGAGCGGTGCGGCCGAC 93
Db	431 AGTATCTGCCCCCAGATTTCGCGCGGCGGTGACAGGTGCGGTGAGAGCCCCATCTGCCCCGTG 372
Qy	94 GCGCAGAACTGGACACGCGAAGGCGCACCCGCGCTGAGACAGACGACGAGGCGGTAGCCG 153
Db	371 GCGCAGAAAGACGACGCGCCATATGCGCCGACGCGCGCTGCGTGTCTGATGTCACACCGTGGCGCG 312
Qy	154 TCGTGGACACGAGATCCGACCGTCGTGATGAAATTGCGCGGTGACCAAGCTCGAACAGGAAC 213
Db	311 CCTTGAAACGCATATGACGACGGTTTCATATGATGTGACCGGTGTCAGTGTGACGACGACGAGC 252
Qy	214 TTGTGCTCTGGCTTCCCGCTGGT 237
Db	251 TTGTGCTGAAAGCGGTGCTGCTG 228

LOCUS	AR619704	13751 bp	DNA	linear	PAT 14-FEB-2005
DEFINITION	Sequence 1099 from patent US 6833447.				
ACCESSION	AR619704				
VERSION	AR619704.1	GI:5924110			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 13751)				
TITLE	Goldman,B.S., Hinkle,G.J., Slater,S.C. and Wiegand,R.C.				
JOURNAL	Mycobacterium xanthus genome sequences and uses thereof				
	Patent: US 6833447-A 1099 21-DEC-2004;				
	Monsanto Technology, LLC; St. Louis, MO				
FEATURES	Location/Qualifiers				
source	1..13751				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
	13.0%; Score 77.6; DB 6; Length 13751;				
Query Match	Best Local Similarity 61.3%; Pred. No. 0.0017;				
Matches 125; Conservative	0; Mismatches 79; Indels 0; Gaps 0;				
QY	34	AACACCTGTGTGACGATCGCCGCGGTGACAGGTGCTCTTAGAGCGCTCGCGGCCGAC	93		
DB	5590	AATTATCTGCCCAACATTTTCGCCGAGCGTCAGGTGGCTGAGAGCCCATCTGCCCGGTG	5533		
QY	94	GCGCAAACTGACACGCGAAGCGCACCCCGCTTGAGA GAGACGACGAGCGGTGTAGCGG	153		
DB	5530	GCGCAGAAAGACGCAACCCATGCGGACGCGCGCTCGATGATGACACACGTTGCGCGG	5471		
QY	154	TGCTGCGACGCGATCGCACCGCTTTCATGATGAATTGCGGTGACACCACTCTGAACAGAAC	213		
DB	5470	CCCTTGAAACCGATGACACGCTTTCATATGATCTGCCGCTCTCATGTCGACGACGAGC	5411		
QY	214	TTTGTGCTGTGCGCTTCCCTGGTG	237		

Db 5410 TTGTCCGTGAAGCCCTCGCTG 5387

RESULT 14
BA000030_32
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragement Name	Begin	End
BA000030_00	1	11000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000
BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000

BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000
BA000030_72	7200001	7310000
BA000030_73	7300001	7410000
BA000030_74	7400001	7510000
BA000030_75	7500001	7610000
BA000030_76	7600001	7710000
BA000030_77	7700001	7810000
BA000030_78	7800001	7910000
BA000030_79	7900001	8010000
BA000030_80	8000001	8110000
BA000030_81	8100001	8210000
BA000030_82	8200001	8310000
BA000030_83	8300001	8410000
BA000030_84	8400001	8510000
BA000030_85	8500001	8610000
BA000030_86	8600001	8710000
BA000030_87	8700001	8810000
BA000030_88	8800001	8910000
BA000030_89	8900001	9010000
BA000030_90	9000001	9025608

Query Match	12.9%	Score 77.2;	DB 1;	Length 110000;
-------------	-------	-------------	-------	----------------

OY	14	TGAGCGCCGGGACCCGAGCGGCAGCAGCACTGTGCGACCATCTTCGCCGCGTATGACAAGTGTCGT	73
Dd	21185	TGAGCGCCCGCACATGCCGTCCACGATCTGGTAGCAGATCTCGCGCGGTGCAGACAGTTCCGGT	21244
OY	74	TGAGCGCCGTGCGCGGCCCGAACGCGCGACAACTGGCACCAGAGGCGCACCCCGGCTGAGAAG	133
Dd	21245	CGAAGCCCCGCTGTGTCGGGTGGGCAAGACGGGGAGATTCAATACCGCACCCGGGCTGGAGAC	21304
OY	134	AGACGACGAGCGCGGTGTAGCCGTGCTGTGGCGACCGATCCGACACCGTCTTCGATGAATAATGGCGT	193
Dd	21305	TGATGCAACATGTCATCCACCCGGTCCGGGTAGCGCATGAGCACCGACTTCGACGACGATCCCGT	21364
OY	194	CGACCAAGCTCGAACAGAGAACTT	215
Dd	21365	CGAACACGCGCCGACACGCTTT	21386
RESULT 15			
AE001946/c			
LOCUS	AE001946	10719 bp	DNA linear BCT 22-NOV-1999
DEFINITION	Deinococcus radiodurans RI section 83 of 223 of the complete chromosome 1.		
ACCESSION	AE001946	AE000513	
VERSION	AE001946.1	GI:6458655	
KEYWORDS			
SOURCE			
ORGANISM	Deinococcus radiodurans RI		
REFERENCE	Deinococcus radiodurans RI		
AUTHORS	Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales; Deinococcaceae; Deinococcus.		
	1 (bases 1 to 10719)		
	White,O., Eisen,J.T.A., Heidelberg,J.F., Hickey,B.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Yamathayan,J.U., Lam,P., McDonald,L., Uitterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.		
TITLE	Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI		
JOURNAL	Science 286 (5444), 1571-1577 (1999)		
PUBMED	10567266		
REFERENCE	2 (bases 1 to 10719)		
AUTHORS	White,O., Eisen,J.T.A., Heidelberg,J.F., Hickey,B.K., Peterson,J.D.,		

Dojson, R. J., Haft, D. H., Gwin, M. L., Nelson, M. C., Richardson, D. L., Moffett, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Yamachew, J. J., Lam, P., McDonald, L., Unterback, T., Zalewski, C., Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W., Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S., Smith, H. O., Venter, J. C. and Fraser, C. M.

TITLE
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
JOURNAL
Medical Center Dr, Rockville, MD 20850, USA
source
Location/Qualifiers
1. .10719
/organism="Deinococcus radiodurans R1"
/mol_type="genomic DNA"
/strain="R1"
/db_xref="taxon:243230"
/chromosome="1"
72. .1034
/gene="DR0932"
72. .1034
/gene="DR0932"
/note="similar to SP:P19641 GB:X68873 PID:388220
PID:606125 GB:U00096 percent identity: 57.96; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polyprenyl synthase"
/protein_id="AAP10509.1"
/db_xref="GI:6458656"
/translation="MPDAAFETMRREVRSDVEFIATIGEDLVTAQGRVPLVTLIA
AQAAGAPDALATSHSWDBLRLSVCVELHSABLDHDLIDSDKRRGQSTARRRG
NVVMSGDDFTLRLGELSGMPSGSPALTRAFQQAAYICGEVLCQVVASDYSE
NYPVITGKTRALIELIAQAAPAMLGADDSRDLSTGREGYMAFQMODDLILSE
EAQIGKVGCDLRGKATPYCLCGPHDAEVYAIERRAGEDVARMIALAQRG
TDDCTREIRRRARLATAALDAPSPARTALAAERELKRTT"
1109. .3043
/gene="DR0933"
1109. .3043
/gene="DR0933"
/note="similar to PID:2094767 percent identity: 60.15;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-amylase"
/protein_id="AAP10510.1"
/db_xref="GI:6458657"
/translation="MLTPDIAARLRKLFDDDDRAETFRRLERYGPELIADNRAYVGN
HADALIGELRVMHAHYARPADLRKLEALRLPRDLQGEWGYAVYDRFAGTLR
GVGRLEYLGGVYVLIHLPLRPDGENGGYAVQDYRSVRPDLITDLSLARE
LRGIGISLVLDLVNVAEHEHMAVRATAGAAVADYFHI PPRTOPDAYERTLPEIF
PDPAQNTNGGAGGWTTTPNRSQMDVMGNAPVPRRYDLITLNRGVEVPRLD
AIAPIKRLIGTDCQNPVHRLTHALRAATTVAVAVAFKAAIYARBDLIHYGSD
HHRGSDMAVNSLNVOLMSSLASRDTRLTALAAAPFKPTNTWGVYVCHDIDG
AIADEDAARVGI.GPARRHFLSDYSGEPGSGFAGLVGHHPTQDRIISGTAASLA
GLDALLETGDAERVDALARLLLAHVMGFGVALLTMGDELALANDTAAVAPHA
ADNRVHRPKQMDLVAQAQDAATGCVTPAGRMFAGLRHLAVRRRTPIHLASTES
RPLSPRCVTLIRREHPTGVLTQVYNSHHITPTPRLQEOUGAVAHDLIGSSQFI
LGEPDLAEFRALMLVAGS"
3065. .3838
/gene="DR0934"
3065. .3838
/gene="DR0934"
/note="similar to GB:AE000657 percent identity: 72.64;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MerR family"
/protein_id="AAP10511.1"
/db_xref="GI:6458658"
/translation="MLVVCSTSLMGSDAKRPVYIVSAAELVDMHPOTLRYEKGL
IRRSGSKTRLYSRDI.EHLREIRLLQELGVNAGVEVNRLLQHODDDQREAE
IERIDELRQARPRALPAPDAPDRPRPVVISIAELVDMHPOTLRYEKGLIH
PQSGKTRLYSERDIEHLREIRRLTOELGVNAGVEIEMRLRHRLDASRAHLEGNVR
RLQDLSRMTTWRLLPAPESAGPEEDAAAAADABDN"

gene

CDS

3816. .4907
/gene="DR0935"
3816. .4907
/gene="DR0935"
/note="similar to GB:D10483 SP:P05637 GB:X04711 PID:216474
PID:40919 percent identity: 55.46; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphatase, putative"
/protein_id="AAP10512.1"
/db_xref="GI:6458659"
/translation="MPEKTTTHERQELSLFRGVRLIREPRFVSDLSRPPAAFPVL
CPKPLNINAAFLPSVPAPRLVWGVGHGANDRLRLLEAGITLIGHMTGGAHL
VPLGDVYDRGDLGVYRLVQSLBEOALAGHAKTALIGHREVLAAALFRQDPPD
RSGFPYWRNSGQTLDERLAPDGLAMRARPLAAQGMILLHADSPLYLGNDP
QEVNRYAEILSPDRPMAAPANAAPDRKAPAGDDOTAAQRLTYGGERLVHGH
PTPLLQSGQDLPKGPVPLVADGICLVADGMAFFDAGFVARLGEIAPGRABR
VLAELVTLPTFLDLSLTATALRPSGEG"
4977. .6017
/gene="DR0936"
4977. .6017
/gene="DR0936"
/note="similar to GB:AU09126 percent identity: 61.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP10513.1"
/db_xref="GI:6458660"
/translation="MPLLDLIDHPHYLBEGRKQLLEWYFVQGVGFDMATNIPAGL
RABLESYQNPFRDITVYSHDSVYLLTLNDGRQMEVYMPYLRKTCVSTWVG
CPARCAFCATGPKGKGNLTGSEIVQGVLLVAGGEGISPEIRNLVIMGEBALLNE
NSMAARVLLHPALGMSKRYLTSTGIAKIRRLAEBDELKIAISLHAPEEER
QQIIPTAANSIEETMAAADYQAVTGRVYTMETMGVNDHLMQALADRLDGLV
SHVNLIPMNPMWDSDFVSSBQIQATFDALQDRGVDSVRSRKGAGAACGLALR
RPGARTGEVGA"
6345. .8135
/gene="DR0937"
6345. .8135
/gene="DR0937"
/note="similar to PID:1652192 percent identity: 41.71;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tetra-ricopeptide repeat family protein"
/protein_id="AAP10514.1"
/db_xref="GI:6458661"
/translation="MLVSPDPRFLAARLLGAWLACGAGAQAQTLDTSAIIVQNT
LNOIQGVTVGSLISVQFDCAKAPSCXQGPVGAAPATSGQSAQNAQSAQAP
SOVAPAAPTPTOTTPPAGEAPATRPVPTVPLTAEQQLQAAQAQFAAGRYQAR
NEERALLVRYNNEPPEHFGIALYQDGLDRAFFRQGFQVQPORREGYNGVIA
TKERYPDALRYEALITLAQSAAPPAARQVILBALATGCGVQVYVLAAYALRA
DPEVETI FRIQAQLYQAGVADLPTVAVLERKPSLQALILADILYVAGCPDRA
VRELSAAARVLTQIDRATLLTQSDLLAQNLRGALNAASEARQEDRRAPAAITRE
ABLLAQGQRAAVRAYADALATPEKSAIQAAALYALBNQYBEAQAQAQALKT
PDASTRALVVGAVVAVROGYQARVLSVLAADQSLIMLGISYALADYAS
AVPILSIVRLRPTPARONLAAALATARVABALLRGVLTVEKNAEGMYLLGLS
ORAQQRPDARQSFRTAAGLSABAQALK"
8132. .9319
/gene="DR0938"
8132. .9319
/gene="DR0938"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP10515.1"
/db_xref="GI:6458663"
/translation="MTRAKTSPGRRPSPRRRPAQRRWPELVATLVALGVGLG
LILDERRAPAPVGAQPTTISIPVSGAVVADGEGTAAAPQAQGTAKIRPAPAP
APKLPPLPBRPAPAPTPPRPVEETTAASPTTQPVQAPATQTPPTQTPAQA
PATQPPATPAPAPAPAPATPTEPATPTEPAPAPAGTAVASQRTPLRSDYV

gene

CDS

gene
CDS

MLGTGSEALRSATAGVSNLGYTVVAIDLNQFVAQVGPFADEASGOQAAADIRRAY
ARAEIYPRGQTLRNPAFAAAEAPATPEBASBPAPTAPETPAAPAPAGPVVYOGA
FNNVGAQRFEQLEBRAQSPSPVNAPEYTKVTVLLGLTNSDLTSTGRDLAAGLDHP
RLR"

9316. .10020
/gene="DR0939"
9316. .10020
/gene="DR0939"

/note="similar to GB:AL009126 percent identity: 51.93;
identified by sequence similarity; putative"

/codon_start=1
/transl_table=11

/product="conserved hypothetical protein"
/protein_id="AAFI0515.1"

/db_xref="GI:6458662"

/translation="MSRVADIPRTATVGRVVTYIRVLEBEAQVNLRASSGELARRAGV
TPQVRKDLTYRGRFRTGIGTVAAVLRRELLRALGIDQTMNVIVIGKRLGHALIANY
PGASDYQFQNVGLFDVAQPDVGRERVRLTIQHSQDLPFVASVAGTPRQVDMGLTVP
ABHAQAAQALVAVAGVGLINFAPVVLQTDLHLPRAFAPAGRRBVTVENVDPLAGMK
RLAFVMLGPOAGPABEE"

10017. .10643
/gene="DR0940"

10017. .10643
/gene="DR0940"

/note="identified by Glimmer2; putative"

/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="AAFI0517.1"

/db_xref="GI:6458664"

/translation="MTMKQWFPVTLILGILLAGCTVGLSGRVNVGVVDVGDGSEOV
ATLITPBEKCDKGVCTPQKQELSTIDGQPVTFPTFARPSSEAVTLEGYRVLSDRLDG

Query Match

12.7%; Score 76; DB 1; Length 10719;

Best Local Similarity 60.8%; Pred. No. 0.0032;

Matches 124; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy	17	CGCCGGGCGCCGAGCGCAGACCTGTCGACGATCGCCCGGTAGCAGGTGCGCTTGA	76
Db	5385	CGCCCCCGGCGACCGCGCAGCAGCCTGCCCGCGATTCGCGGGGTGAGGTGCGCCGA	5326
Qy	77	GGCCGTGCGGCGCCGCGCAGCGCAGACTGCGACGCGAAGCGCACCCCGCTGAGACGAGA	136
Db	5325	AGCCCATCTTGCCTGCGGCGCAAAAGCGGCGGCGCGGCGCAACCCACATGTCGAGA	5266
Qy	137	CGCAGCGCGTGTAGCCGTGCTGCGCAGATCCGACCGCTTCGATGAATTCGGTCCA	196
Db	5265	CGCAGATGCTCTTGCCTGCGGTGAGGATGTACACCGCTTCATCTGCCGCCGCTCGT	5206
Qy	197	CCAGCTCGAACAGAACTTTGTG	220
Db	5205	TGAGGCTGAACAGATTTTGACCG	5182

Search completed: April 7, 2006, 16:57:46
Job time : 2893.48 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:41:57 ; Search time 226.527 Seconds
(without alignments)
10613.880 Million cell updates/sec

Title: US-10-611-442-2_COPY_1200_1800

Perfect score: 601

Sequence: 1 ggcctcctacgcggcg99gcaaa.....ggacgcgcctacgcgcact 601

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 9267905 seqs, 2000278028 residues

Total number of hits satisfying chosen parameters: 18535810

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:*

1: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*

2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*

3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*

4: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*

5: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*

6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*

7: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*

8: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*

9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*

10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*

11: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*

12: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*

13: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*

14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*

15: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174.6	29.1	116856	14	US-11-143-980-1
2	150.4	25.0	6210	14	US-11-056-470-1
3	132	22.0	6360	14	US-11-056-470-2
4	101.2	16.8	10524	14	US-11-075-185-37
5	101.2	16.8	78869	14	US-11-075-185-1
6	97.2	16.2	42393	14	US-11-175-689-3
7	96.2	16.0	11070	14	US-11-075-185-34
8	95.6	15.9	50973	14	US-11-175-689-1
9	91.2	15.2	10968	14	US-11-075-185-35
10	71.8	11.9	37507	8	US-10-522-037-2
11	70.8	11.8	3381	14	US-11-075-185-33
12	70.2	11.7	1694	11	US-11-254-686-11
13	69.6	11.6	7233	14	US-11-175-689-2
14	68.4	11.4	444	8	US-10-802-796-658
15	68.4	11.4	444	11	US-11-221-284-658
16	68	11.3	367	8	US-10-802-796-627
17	68	11.3	367	11	US-11-221-284-627
18	62.8	10.4	5679	14	US-11-075-185-36

19	61.6	10.2	1536	11	US-11-096-568A-22690	Sequence 22690, A
20	57	9.5	1619	11	US-11-096-568A-15979	Sequence 15979, A
21	55.6	9.3	88421	14	US-11-205-109-1	Sequence 1, Appl1
22	54.4	9.1	5301	14	US-11-075-185-40	Sequence 40, Appl1
23	53.6	8.9	1601	11	US-11-096-568A-17412	Sequence 17412, A
24	52.8	8.8	1612	14	US-11-129-143-157	Sequence 157, App
25	52.2	8.7	1854	11	US-11-096-568A-13341	Sequence 13341, A
26	52	8.7	1759	11	US-11-096-568A-21990	Sequence 21990, A
27	51.4	8.6	7364	14	US-11-096-281-10	Sequence 10, Appl1
28	51.2	8.5	1473	14	US-11-110-082-8	Sequence 8, Appl1
29	51.2	8.5	37507	8	US-10-522-037-2	Sequence 2, Appl1
30	50.8	8.5	2743	8	US-10-517-939-359	Sequence 359, App
31	50.6	8.4	6413	11	US-11-134-587B-11	Sequence 11, Appl1
32	50.6	8.4	6604	11	US-11-134-587B-11	Sequence 11000, A
33	50.2	8.4	1326	11	US-11-096-568A-11000	Sequence 21514, A
34	49.8	8.3	1426	11	US-11-096-568A-21514	Sequence 26800, A
35	49.2	8.2	1685	11	US-11-254-686-8	Sequence 8, Appl1
36	49	8.2	931	11	US-11-096-568A-26800	Sequence 20778, A
37	48.8	8.1	1607	11	US-11-096-568A-20778	Sequence 1, Appl1
38	48.4	8.1	88421	14	US-11-205-109-1	Sequence 104, App
39	48.2	8.0	1242	14	US-11-143-401-104	Sequence 16072, A
40	47.8	8.0	955	11	US-11-096-568A-16072	Sequence 347, App
41	47.6	7.9	902	9	US-10-714-887-347	Sequence 19208, A
42	47.6	7.9	1409	11	US-11-096-568A-19208	Sequence 12992, A
43	47.4	7.9	1292	11	US-11-096-568A-12992	Sequence 21138, A
44	47.2	7.9	862	11	US-11-096-568A-21138	Sequence 10705, A
45	47.2	7.9	1828	11	US-11-096-568A-10705	

ALIGNMENTS

RESULT 1

US-11-143-980-1

Sequence 1, Application US/11143980

Publication No. US20050272133A1

GENERAL INFORMATION:

APPLICANT: He, Min

APPLICANT: Hecul, John

APPLICANT: Haltli, Bradley A.

APPLICANT: Wagnenaar, Melissa M.

APPLICANT: Graziani, Edmund

APPLICANT: Summers, Ma

APPLICANT: Kulowski, Kerry

APPLICANT: Peng, Kevin

TITLE OF INVENTION: Biotynthetic Gene Cluster for the Production of a Complex

TITLE OF INVENTION: Polyketide

FILE REFERENCE: AM-101426US

CURRENT APPLICATION NUMBER: US/11/143,980

CURRENT FILING DATE: 2005-06-03

PRIOR APPLICATION NUMBER: US 60/664,483

PRIOR FILING DATE: 2005-03-23

PRIOR APPLICATION NUMBER: US 60/576,895

PRIOR FILING DATE: 2004-06-03

NUMBER OF SEQ ID NOS: 72

SOFTWARE: PatentIn version 3.3

SEQ ID NO 1

LENGTH: 116856

TYPE: DNA

ORGANISM: Streptomyces sp.

US-11-143-980-1

Query Match 29.1%; Score 174.6; DB 14; Length 116856;
Best Local Similarity 57.7%; Pred. No. 1.5e-29;
Matches 349; Conservative 0; Mismatches 229; Indels 27; Gaps 1;

QY 1 ggcctcctacgcggcg99gcaaa.....ggacgcgcctacgcgcact 60

DB 52600 ggcaactacgcacccgcgcactgtctgacacactgacgagcggcgccgag 52659

QY 61 ggcctcctacgcggcg99gcaaa.....ggacgcgcctacgcgcact 120

DB 52660 ggcctcctacgcggcg99gcaaa.....ggacgcgcctacgcgcact 52719

;; PRIOR FILING DATE: 2002-07-12
;; NUMBER OF SEQ ID NOS: 10
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2
;; LENGTH: 6360
;; TYPE: DNA
;; ORGANISM: Streptomyces geldanus
US-11-056-470-2

Query Match 22.0%; Score 132; DB 14; Length 6360;
Best Local Similarity 55.2%; Pred. No. 4,6e-20;
Matches 327; Conservative 0; Mismatches 250; Indels 15; Gaps 3;

QY 1 GCGCTCCAGCGGCGGCAAGCGCTTCTCGACGCGCTTCCGCGCGCTGCGGCGAGT 60
DB 5656 GCGACACTACCTCCCGCCAAAGCTTATCTGAGCGACTTCGACAGTGGCGGCTGCGCAC 5715
QY 61 GGGCTGCGGCTCACTTCGATCGCTGCGGCTCTGTGGCCGCGGCAAGAA---CATGGCCGCT 117
DB 5716 GGGCTCCCGCACTCGCTGCGGCGATGGCGATGGCGGCAAGCAAGTGGCATGAGCGCC 5775
QY 118 ACCGAGGCGGCGCACTACCTGCGGCGAGCGGCTGCGGCGCATGAGACCGGCGGCG 177
DB 5776 GGGCTGCGGCTCGGCGGCTGCGGCGAGTGGCAAGCGGCTGCGGCGCATGCTGCGGCTGCGACG 5835
QY 178 ATGAGAGGCTGCGGCAAGCAAGCGGCGGAGCGGCGGAGCGGCGGCTGCGGCGGCTGCGG 237
DB 5836 GCGGAGGCGCTGAGCGCTTTCGACCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 5895
QY 238 GACCGGAGCGGCTTCTGCGGCTTTCGACCGGCGGCGCGCGCGCGCGCGCTTTCGAGCA 296
DB 5896 TTGCGGCTGCGGCTTTCG 5955
QY 297 ACTCGGCTGCGGCTTTCG 345
DB 5956 GGGCTGCGGCTTTCG 6015
QY 346 CTGCGCGCGCGCGGCTGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 405
DB 6016 CTGCGCGCGCGCGGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6075
QY 406 GTTCGAGCGGCGGCTGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 465
DB 6076 ATCCGAGCGGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 6135
QY 466 GTTCGCGGCTTTCGAGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 525
DB 6136 CAGGCGGCTTTCGAGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6195
QY 526 GCGGCGGCTTTCGAGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577
DB 6196 AACACGCGCGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6247

RESULT 4

US-11-075-185-37
;; Sequence 37, Application US/11075185
;; Publication No. US20050266434A1
;; GENERAL INFORMATION:
;; APPLICANT: REEVES, CHRISTOPHER D
;; APPLICANT: JULIEN, BRYAN
;; APPLICANT: REID, RALPH
;; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBUTICINS
;; FILE REFERENCE: 010099.03
;; CURRENT APPLICATION NUMBER: US/11/075,185
;; PRIOR FILING DATE: 2005-03-07
;; PRIOR APPLICATION NUMBER: US 60/551,103
;; PRIOR FILING DATE: 2004-03-08
;; PRIOR APPLICATION NUMBER: US 60/568,290
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 37

;; LENGTH: 10524
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-11-075-185-37

Query Match 16.8%; Score 101.2; DB 14; Length 10524;
Best Local Similarity 59.3%; Pred. No. 2e-13;
Matches 172; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 305 GGGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 364
DB 5039 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5098
QY 365 CGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 424
DB 5099 GACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5158
QY 425 CGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 484
DB 5159 GCGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5218
QY 485 GATTGCACTTCATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544
DB 5219 GCGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5278
QY 545 GGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 594
DB 5279 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5328

RESULT 5

US-11-075-185-1
;; Sequence 1, Application US/11075185
;; Publication No. US20050266434A1
;; GENERAL INFORMATION:
;; APPLICANT: REEVES, CHRISTOPHER D
;; APPLICANT: JULIEN, BRYAN
;; APPLICANT: REID, RALPH
;; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBUTICINS
;; FILE REFERENCE: 010099.03
;; CURRENT APPLICATION NUMBER: US/11/075,185
;; PRIOR FILING DATE: 2005-03-07
;; PRIOR APPLICATION NUMBER: US 60/551,103
;; PRIOR FILING DATE: 2004-03-08
;; PRIOR APPLICATION NUMBER: US 60/568,290
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 1
;; LENGTH: 78869
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match 16.8%; Score 101.2; DB 14; Length 78869;
Best Local Similarity 59.3%; Pred. No. 1.4e-13;
Matches 172; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 305 GGGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 364
DB 36945 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 37004
QY 365 CGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 424
DB 37005 GACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 37064
QY 425 CGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 484
DB 37065 GCGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 37124
QY 485 GATTGCACTTCATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544
DB 37125 GCGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 37184

QY 545 GGGTGGCCAGACCATGCTTTCAGACCAAGTGAAGCCGCTGACC 594
Db 37185 AGACGCGCGGAGCGCTGCTTTCAGACCAAGTGAAGCCGCTGACC 37234

RESULT 6

US-11-175-689-3
; Sequence 3, Application US/11175689
; Publication No. US20060024806A1
; GENERAL INFORMATION:
; APPLICANT: STINEAR, TIMOTHY P.
; APPLICANT: COLE, STEWART T.
; APPLICANT: LEADLAY, PETER F.
; APPLICANT: SMALL, PAMELA L.C.
; APPLICANT: JOHNSON, PAUL D.R.
; APPLICANT: JENKIN, GRANT A.
; APPLICANT: DAVIES, JOHN K.
; APPLICANT: HAYDOCK, STEPHEN F.
; TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING
; TITLE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES
; FILE REFERENCE: 03495, 0329-01
; CURRENT APPLICATION NUMBER: US/11/175,689
; PRIORITY FILING DATE: 2005-07-07
; PRIOR APPLICATION NUMBER: 10/987,592
; PRIOR FILING DATE: 2004-11-15
; PRIOR APPLICATION NUMBER: 60/519,864
; PRIOR FILING DATE: 2003-11-14
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 42393
; TYPE: DNA
; ORGANISM: Mycobacterium ulcerans
US-11-175-689-3

Query Match 16.2%; Score 97.2; DB 14; Length 42393;
Best Local Similarity 49.0%; Pred. No. 1,1e-12;
Matches 301; Conservative 0; Mismatches 298; Indels 15; Gaps 1;

QY 1 GCTTCCTACGGCGGCGGAGAGCCCTTCTCGACCGCTTGGCGGTGCTGCGGCGAGT 60
Db 13996 GCGCGATTCGCTGCGCTGACGATCTTGAATTCGCTGCGGAGAAAGCGGACGTTG 14055
QY 61 GGGTGGCGGCTGACCTGATCGCTGCGGCTGCTGCGGCGGAGAGAAAGCGGCGTACC 120
Db 14056 GGGTTCACGAGACATCGCTGCTTGGGAGCTGCGGCTGCGGCGGATGAGAGACCGG 14115
QY 121 GAGGCGGCGGACCTACCTGCGAGCGGCGCTGCGGCGGATGAGACCGGAGCGGCGATC 180
Db 14116 CCATCCGCTTGGGAGCTTAACCTTAATGCTTGAATGATGTCAGCAGATTGAGCTGTG 14175
QY 181 GAGAGCTGCGGACCACTTGGAGCGCGGAGACCGGTGAGTGTGCTGCTGAGACTGAGAC 240
Db 14176 CAGGCGCTGACGAGCGCATTTGACGAGCGGAGACCATTTGACTGTGCGAGCGCTCAAC 14235
QY 241 CGGAGCGGCTTCTGCGAAGCTTTGACCGCGCGCGCGCGCGCGCTTCTTGAAGAACTC 300
Db 14236 TGGGATCGGTTCTACCCCACTTCGCTTTGGCGGAGCGAGCGCTTCTTACAGAAATC 14295
QY 301 GGTGGGTGCGCGCGCGGCGGAGAGACCGGTGAGAGATGATCTGCGCGCGGCGCTG 360
Db 14296 ACGAGGTAATGAGCTTACCGGAGTGTGAGCTGAGACTGTCATGACCGGCGAGCGCTC 14355
QY 361 GCGTGA-----TGCAGAGCGGAGAGCTGACGAGAGATGTCGCGCGCTG 405
Db 14356 CTGACGAGCAATTAAGCGGAGCTTACCGGAGCAGAGACGCTGACGAGTGAAGTTG 14415
QY 406 GTCCGAGCGGAGTGGAGCGGTGCTGGGCGACCGGAGCGGAGCGGAGTGAAGGCTGAC 465
Db 14416 GTCTTGAATCAAGCGGAGCGGTGCTGGGAGCGGCTTCACTGAGAGTCTGATATCTCAT 14475
QY 466 GTCCGCTTCCGTAAGCTGGGATTCGATCTCAATGACCGCGGTGACCTGCGAGAACCGGCTC 525

Db 14476 GAGTCATTCAAAACCTCGGATTTGATTCGCTACCGGCTTGAACCTGCGGACCTTC 14535
QY 526 GCGGCGGTGACCGGCGGTCCGGGTGCGGAGACCATGCTTTCGACCAAGGAGTGAAC 585
Db 14536 CAAACTGCGAGCGGCGCTTCACTTTGCTCCACTTATCTTCAATCAACCCACACCCCAT 14595
QY 586 CGCTTCACCGCGCA 599
Db 14596 GCGGTGCGGAGCA 14609

RESULT 7

US-11-075-185-34
; Sequence 34, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; PRIORITY FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
; LENGTH: 11070
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-34

Query Match 16.0%; Score 96.2; DB 14; Length 11070;
Best Local Similarity 49.7%; Pred. No. 2,4e-12;
Matches 271; Conservative 0; Mismatches 273; Indels 1; Gaps 1;

QY 34 GCTTCGCGCGCTGCGGCGGAGGCGGCTGCGGCGGATGACCTGATGCTGAGGCTG 93
Db 3914 GCTTCGCGCGGCTTCTGAGAGAGCGCGGCTGCGGATGAGCGGACCTTCCGCGCG 3973
QY 94 TGGGCGGCGGAGACATGCGCGGTACCGAGCGGCGGAGCTGCTGCGGAGCGGCGCTG 153
Db 3974 CCGACTGCGGAGCGCTGCGGAGCGGAGCGGCGCTTGTGCGGCTGCTGCTGCGGAGCGG 4032
QY 154 CGGCGCATGAGCGGCGGCGGCGGATGAGGAGCTGCGGAGCACCTTGAACGCGGAGAC 213
Db 4033 CGCTCTCGGAGCGGCGGCGGCTGCGGAGCGGAGCGGAGCGGCTGCGGCGGCTGCGAC 4092
QY 214 CCGTGGGTGCTGCGGAGCTGAGACCGGAGAGGCTTCTGCGAAGCTTTCACCGCGCGC 273
Db 4093 CTGTGCGGCTTCCAGCGGTGCGGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 4152
QY 274 CGGCGCGGCGGCTTCTTCAAGAACTGCGGTGCGGCGGCGGCGGCGGAGAGAGAGCGGT 333
Db 4153 CTGTGCGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGAGAGAGAGAGAGAGAGAGCG 4212
QY 334 CAGGAACTGATCTTCCGCGGCGGCTGCGGAGTGTGAGAGCGGAGAGAGAGAGAGAGAT 393
Db 4213 GCGCGGTGCGGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4272
QY 394 GTCCGCGGCGGCTGCGGAG 453
Db 4273 CTGTCTGAGCTGCGGAG 4332
QY 454 ATTCAGAGTGAAGTTCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 513
Db 4333 GTCCCGGAG 4392
QY 514 CCGAAGCGGCTGCGGCGGAG 573

[illegible]

```

RESULT 10
US-10-522-037-2
; Sequence 2, Application US/10522037
; Publication No. US20050282166A1
; GENERAL INFORMATION:
; APPLICANT: LIBRAGEN
; TITLE OF INVENTION: Method for the expression of unknown environmental DNA into adapt
; TITLE OF INVENTION: cells.
; FILE REFERENCE: B0149M0
; CURRENT APPLICATION NUMBER: US/10/522.037
; CURRENT FILING DATE: 2005-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO. 2
; LENGTH: 37507
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: DNA sequence of clone FS3-135.
US-10-522-037-2

```

Query Match	11.9%	Score 71.8	DB 8	Length 37507
Best Local Similarity	48.4%	Pred. No. 3.8e-07		
Matches 299; Conservative	0	Mismatches 302	Indels 12	Gaps 3

QY 1 GCGCTCTACGGGGGGGAGAGCGCTTCCTCGACGCCCTTCGACCCTGCGTGTGTCCGGCCAGT 60
 Db 13270 GCCAACTACATCGCGGCCCACTGGCTTCTCTCGAGGACACTCCGCCAAGGGCGCCCGTCGTGAG 13322
 QY 61 GGGCTGCGCGGTCACTTCGATGCGCTTGGG---TCTGTGGGCGGGGAGAACTATGCG--- 113
 Db 13330 GGGCCAGCGCGCGCTGTGTGGGTGGGCTGGGGCGCGATCGGGGAGCGTGGGCTACTTCGCGCGCG 13388
 QY 114 --CGGTACCGAAGGGCGCGGACTTACTCTGCGCAGCCAGCGGCGCTGCGCGCGCTATGACCCGCGAG 171
 Db 13390 CACGAGGAAGGTGAAGAGAAAGCCCTCGACAGCGCACTGGGGCGGACACCGGACTTCGATGGGAG 13445
 QY 172 CGGGGAGATCGAGAGAGCTGCGGAGCACACCTTGAACGCGCGGGGAGCCCGTGGGTGTGGTGGTGTG 231
 Db 13450 GCGCGGCTCGCGGTGTCTGAGAGAGCTCTGTCTCGCCACAGCGCTTGGGCGCTTCGGCGTGTCTC 13508
 QY 232 GACCTGGAACCGGAGACGGGTTCGTGAACTGTGTCAACGCGCGCGCGCGCGCGCGCCCTCTTTC 291
 Db 13510 GACTTCAGCTGGCGCGAACCTCGCGCGGCTTCTCTCCACACGCGCGCGCTCGCGCGCGCTTCGCG 13565
 QY 292 GACGAACCTCGGTGGGGGTTCGCGCGCGCGGGGCGGAGAGACCGGTCAAGGAATGGATCTCGCC 351
 Db 13570 GAGCTGAGCGCGCGCGGCGAGGACGCGGAGAGACGACGACCGCGCGCGCGCGAGCTCGCG 13622
 QY 352 CGGCGGCTGCGGTGATGCGCGGAGGCGGAAAGTCAAGAGCATGTTCGCGCGGCTTGCTGCGA 411
 Db 13630 CGCTTCGCGCAAGAGACTGAGAGCGCTCCGAGCTTCCTGGGAGATGTTACCACTGACTGTCTGCGC 13688
 QY 412 GCGGAGGTGACAGCGGTCTCTGAGGCACCGGCACCGCGGACCGGTATCGAGCGTGAACGTTCGCC 471
 Db 13690 CGCGAGGTGGCGGAGATCTCG---CGATTCGCGCGGACCGCGCTTCGATTAAGCGCGCAGGCC 13744
 QY 472 TTCCTGACTGGGAATCGACTTCATGACCGCGCTCGACTTGGCGAAACCGGCTCGCGCGG 531

Db 13747 TTGAGGAAATN96SCATGATTGCTGATGAGGCGTGAAGCTGTGACCCCGAGGAGGCG 13806
QY 532 GTAAACCGGGGTGTCGGGGTGGCAGACCAATGTTTTCACACCCGACATGGACCGGCTTC 591
Db 13807 CGCTTCGGGCGTGAACCTGCGGTGATGTGCTCTCCGACAGCCGTCGATCGAAGAGCTG 13866
QY 592 ACCGGCGAC 600
Db 13867 GTTCGACGCG 13875

```

RESULT 11
US-11-075-185-33
Sequence 33 Application US/11075185
Publication NO. US20050266434A1
GENERAL INFORMATION:
APPLICANT: REEVES, CHRISTOPHER D
APPLICANT: JULIEN, BRYAN
APPLICANT: REID, RALPH
TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
FILE REFERENCE: 010099.03
CURRENT APPLICATION NUMBER: US/11/075,185
CURRENT FILING DATE: 2005-03-07
PRIOR APPLICATION NUMBER: US 60/551,103
PRIOR FILING DATE: 2004-03-08
PRIOR APPLICATION NUMBER: US 60/568,290
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.3
SEQ ID NO 33
LENGTH: 3381
TYPES: DNA
ORGANISM: Sorangium cellulosum
US-11-075-185-33

```

Query Match	11.84;	Score 70.8;	DB 14;	Length 3381;
Best Local Similarity	48.84;	Pred. No. 9.8e-07;		
Matches 251; Conservative	0;	Mismatches 257;	Indels 6;	Gaps 2;

QY 68 CGGTCACTTCATGCGCTTGCGGGTCTGTGGGCGCGGCGAGAACATGAGCGGTATCCGAGAGGCG 127
 Db 2687 CGCAACCGGGTGTCTGGCGGTCCGCGCTGACCGCGGGGTGACCGCGAGCGCGAGAGGCGGTGGG 2746
 QY 128 GCGACTACCTGCGGACGCCAGGGCTTGCGCGCATGGAACCCGCGAGCGGGCGATTGAGAGAC 187
 Db 2747 TCGGAGGCTTGCAACCGGACGACGCGCGGGCTTGGCGGCTGTGCTGGGCGCGCTGGGCGCGC 2806
 QY 188 TCGCGGACCAACCTTGAGACGCGCGGGGACCCGTTGGGGGTGTGGGTGTGGAACCTTGAGACGGGAGC 247
 Db 2807 TGCATGTGCGAGGGCGAGCGCGGTGCATTGGGCGCGGGTCTTGGCGCGCGCGCTTGGCGCGCGCGC 2866
 QY 248 GATTTCGTGAACTGTTCAACGCGCGCGCGCGCGCGCGCGCCCTCTTTCAGAGAACTTGCGTGGG 307
 Db 2867 TGTGTGAACCTGTCCGACGTAAGCAATTTCCAGGCGCGACGGCTACTGTGGTTGAATACCGACGAGA 2926
 QY 308 TCCGCGCGCGGGGCGGAGGAGACCGG---TCAGGAATGTGAATCTTCGCCCGCGGCTGGCGT 354
 Db 2927 GCGTTTGCCTTCGACAGCGCGCTGCCAGATTGGAGAGAGTGTGGATCAGAAATCTGGCCG 2986
 QY 365 CGATGCCCGGAAGCGGAAAGTTCAGAGAGATGTCCGCGCGCTGGTCCGAGCGGAGGTGGCAG 424
 Db 2987 GCGTGTCTCCCGCGCGAGCGGAGAAACGGCTCTCGAATGGGTGTCCGCGCAGAAATCCCG 3046
 QY 425 CGGTGCTGGGACACGGGCAACGCGCGAGCGGTGATTCGAGCTGACGTTCGCTTCCGTGAACCTGG 484
 Db 3047 CGGTGTCTGGGCTGAGAGCGCGCGCGCGG---CCGCGGAGAGTCCCGCTGCGGGAGTCTCG 3103
 QY 485 GATTGACTCATGACCGCGGTGCAACTGTGCGAAACCGGACTTCGCGGCGGTGACCGGGGTCC 544
 Db 3104 GGTGAATTCGCCGATTCGCGTGGAGCTGGGGAGACCGCTTGGAGCGCAGAGACACGCGGGA 3163
 QY 545 GGGTGGCCACGACCATGTCTTTCGACCACTCCGAC 578

```

FILE REFERENCE: 03495_0329-01
CURRENT APPLICATION NUMBER: US/11/175,689
CURRENT FILING DATE: 2005-07-07
PRIOR APPLICATION NUMBER: 10/987,592
PRIOR FILING DATE: 2004-11-15
PRIOR APPLICATION NUMBER: 60/519,864
PRIOR FILING DATE: 2003-11-14
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 3.3
SEQ ID NO 2
LENGTH: 7233
TYPE: DNA
ORGANISM: Mycobacterium ulcerans
US-11-175-689-2

Query Match      11.6% Score 69.6; DB 14; Length 7233;
Best Local Similarity 52.8%; Pred. No. 1.5e-06;
Matches 150; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 317 GGGCGGAGAGACCGGCTCAGAAATCGATCTCGCCGCGAGCTGCGTCGATCCGAGG 376
Db 6110 GGGCGCGGGGGGCGCTCGCGCAACGATTCGCTGCGCCCTCAACGACTTAGCCCC 6169
Qy 377 CCGAAGCTCAGAGCATGTGCGCCGGCTGCTCGAGCCGAGGTGCGACGGCTGTGGCC 436
Db 6170 AACGCAACAAACAAACACTGCGCCACCTCGTGGCCGCGCCACCGCCACCGTCTGGGCC 6229
Qy 437 ACGGCAACGCGGAGGATGATGAGAGGTGAGCGTGGCTTCCGTGACTGGATTGACTCCA 496
Db 6230 AACACACCCCCGAAAGCATAGCCGACGCAACCGCTTCAAGACTTCGGAATGATTCGCG 6289
Qy 497 TGACCGCGCTCGACTTGCAGAACCGGCTCGCGCGGAGTACCGGAGTCCGGGTGCGCAGA 556
Db 6290 TGACCGCGCTTGAACCTGCGGAACCTTCAACCCACCAACCGGCTCGATCTGCCCCCA 6349
Qy 557 CCAATGCTTTCGACCAACCGGACAGTGGACCGCTTCAACCGCGCAC 600
Db 6350 CCTCATCTTCGATCAACCCGACCCGACCCGATCGCTAACCACAC 6393

RESULT 14
US-10-802-796-658/c
Sequence 658, Application US/10802796
Publication No. US20050250104A1
GENERAL INFORMATION:
APPLICANT: COLE, STEWART
APPLICANT: BUCHRIESEN-BROSCH, ROLAND
APPLICANT: GORDON, STEPHEN
APPLICANT: BILLAULT, ALAIN
TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
TITLE OF INVENTION: DNA LIBRARY APPLICATION TO THE DETECTION OF
FILE REFERENCE: 05394_0011-00000
CURRENT APPLICATION NUMBER: US/10/802,796
CURRENT FILING DATE: 2004-03-18
PRIOR APPLICATION NUMBER: US/09/673,476
PRIOR FILING DATE: 2002-03-29
PRIOR APPLICATION NUMBER: PCT/IB99/00740
PRIOR FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 09/060,756
PRIOR FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.2
SEQ ID NO 658
LENGTH: 444
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: modified_base
LOCATION: (296)
OTHER INFORMATION: a, t, c or g
FEATURE:

```

NAME/KEY: modified_base
LOCATION: (314)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (367)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (377)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (394)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (410)
OTHER INFORMATION: a, t, c or g
US-10-802-796-658

Query Match 11.4%; Score 68.4; DB 8; Length 444;
Best Local Similarity 53.1%; Pred. No. 4.8e-06;
Matches 138; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 340 TCGGATCTCGCCCGGCTGGCGTCGATGCCGAGGCCGAACGTCAGAGCATGTGCC 399
DB 412 TCNCGGTTTGGCCGCCGANTGGCCGTCGTCGCGCNCAGCAANTCGCCGACGATC 353
QY 400 CGGCTGGTCCGAGCCGAGGTGCGAGCGGTGCTGGGCCACAGCCGACCGGTATCGAG 459
DB 352 GAGGTGATGTGACATCCGCGACCGGTGCTGGGGCGAGCCGCGCTGCCGACTGAC 293
QY 460 CGTACGTCGCTTCCGTGACCTGGAATTCATCTCACTGACCCGCTGACCTGGGAAAC 519
DB 292 GCTGCCAGCGGCTTGGCCGATTCGGGATTTAATCTTGAGTCCGTGAGATACGTAAAC 233
QY 520 CGGCTCGCGCGGTGACCGGGGTCCGGGTGGCCACGACCATGCTTGAACACCGACA 579
DB 232 CGCTTAACAGCTGTCAACCGAGTAACGCTGCCGGCCACCGGATTCGATCACCCACC 173
QY 580 GTGACCGCTTCAACCGGCA 599
DB 172 CCGACCGAAGTACCCAGTA 153

RESULT 15

US-11-221-284-658/c
Sequence 658, Application US/11221284
Publication No. US2006063182A1
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/11/221,284
CURRENT FILING DATE: 2005-09-08
PRIOR APPLICATION NUMBER: US/09/060,756
PRIOR FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 658
LENGTH: 444
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"

US-11-221-284-658

Query Match 11.4%; Score 68.4; DB 11; Length 444;
Best Local Similarity 53.1%; Pred. No. 4.8e-06;
Matches 138; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 340 TCGGATCTCGCCCGGCTGGCGTCGATGCCGAGGCCGAACGTCAGAGCATGTGCC 399
DB 412 TCNCGGTTTGGCCGCCGANTGGCCGTCGTCGCGCNCAGCAANTCGCCGACGATC 353
QY 400 CGGCTGGTCCGAGCCGAGGTGCGAGCGGTGCTGGGCCACAGCCGACCGGTATCGAG 459
DB 352 GAGGTGATGTGACATCCGCGACCGGTGCTGGGGCGAGCCGCGCTGCCGACTGAC 293
QY 460 CGTACGTCGCTTCCGTGACCTGGAATTCATCTCACTGACCCGCTGACCTGGGAAAC 519
DB 292 GCTGCCAGCGGCTTGGCCGATTCGGGATTTAATCTTGAGTCCGTGAGATACGTAAAC 233
QY 520 CGGCTCGCGCGGTGACCGGGGTCCGGGTGGCCACGACCATGCTTGAACACCGACA 579
DB 232 CGCTTAACAGCTGTCAACCGAGTAACGCTGCCGGCCACCGGATTCGATCACCCACC 173
QY 580 GTGACCGCTTCAACCGGCA 599
DB 172 CCGACCGAAGTACCCAGTA 153

Search completed: April 7, 2006, 03:50:14
Job time : 228.527 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 2314.42 Seconds
(without alignments)
12149.485 Million cell updates/second

```
Title: US-10-611-442-2_COPY_1200_1800
Perfect score: 601
Sequence: 1 gctcctacgcgcyggcaaa.....ggaccgcctcacgcgcact 601
```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 beqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  gb_est1: *
2:  gb_est2: *
3:  gb_est3: *
4:  gb_hc: *
5:  gb_est4: *
6:  gb_est5: *
7:  gb_est6: *
8:  gb_est7: *
9:  gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	74.2	12.3	925	10	CNS0091P	AL053013 Drosophila
C 2	73.6	12.2	932	10	CNS0072Q	AL065742 Drosophila
C 3	72.4	12.0	1088	10	AG075424	AL0675424 Pan trogl
C 4	71.8	11.9	935	10	CNS0065X	AL066051 Drosophila
C 5	70.2	11.7	1033	10	CG156677	CG156677 P051-4-D0
C 6	70	11.6	925	10	CNS0091P	AL053013 Drosophila
C 7	70	11.6	1270	10	AG382572	AG382572 Mus musc
C 8	69.6	11.6	776	10	CNS010RY	AL0993352 Drosophila
C 9	69.6	11.6	1003	5	BO641288	BO641288 AGENCOURT
C 10	68.8	11.4	1007	5	BO650254	BO650254 AGENCOURT
C 11	68.8	11.4	1009	10	CNS010EW	AL098882 Drosophila
C 12	67.8	11.3	886	9	AZ196444	AZ196424 SP 1033 A
C 13	66.4	11.0	908	10	CW949347	CW949347 TCG33.4 G
C 14	66	11.0	1547	10	AG371264	AG371264 Mus musc
C 15	65.4	10.9	907	2	BG946266	BG946266 1024012H1
C 16	65.2	10.8	932	10	CNS0072Q	AL065742 Drosophila
C 17	65	10.8	1151	7	CK210749	CK210749 FGAS02257
C 18	64.8	10.8	996	10	AG072414	AG072414 Pan trogl
C 19	64.8	10.8	998	10	AG081195	AG081195 Pan trogl
C 20	64.6	10.7	1569	10	AG341503	AG341503 Mus musc
C 21	64.2	10.7	1038	10	AG060254	AG060254 Pan trogl
C 22	64.2	10.7	1280	10	CL470317	CL470317 SAIL 140

C	23	64	10.6	977	10	CL507759	SA1L_786
C	24	64	10.6	1552	10	AG430101	Mus muscult
C	25	63.6	10.6	1244	2	BG846745	1024014HO
C	26	63.2	10.5	834	10	AG392512	Mus muscult
C	27	63.2	10.5	1201	10	AG435403	Mus muscult
C	28	63	10.5	894	10	AG075627	Pan trogl
C	29	63	10.5	996	5	BG531181	AG075627
C	30	63	10.5	1101	10	CNS01767Y	AG075627
C	31	63	10.5	1203	10	CNS01554	AG075627
C	32	62.8	10.4	1060	5	BH538693	AL108460
C	33	62.8	10.4	1317	10	AG430654	Drosophill
C	34	62.6	10.4	916	10	CW938719	BH538693
C	35	62.4	10.4	2420	10	CL509247	AG430654
C	36	62.2	10.3	989	5	BH952601	CW938719
C	37	62.2	10.3	1450	10	AG429738	TCB18.4.E
C	38	62	10.3	1359	10	CL507830	SA1L_788
C	39	62	10.3	1661	10	AG435089	Mus muscult
C	40	61.8	10.3	539	6	CA741271	AG435089
C	41	61.8	10.3	839	6	CNS00404B	CA741271
C	42	61.6	10.2	707	6	CA160031	Walc.pko
C	43	61.6	10.2	1157	3	BH542947	AL054280
C	44	61.6	10.2	856	10	AG863445	SC1PR23c0
C	45	61.4	10.2	1039	10	CW918239	AG863445
C	46	61.4	10.2	1039	10	CW918239	Oryza sat
C	47	61.4	10.2	1039	10	CW918239	BDCA5456TRR

ALIGNMENTS

RESULT 1					
CNS0091b/c					
LOCUS	CNS0091P	925 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RP11-38 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

VERSION AL053013.1 GI:4934461

SOURCE *Drosophila melanogaster* (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Ephydroidea; Drosophilidae; Drosophila

Genoscope.
Authors.
Editor.
Director.

JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
DN 101 01006 Error message (3)

- Web : www.genoscope.cns.fr

collaboration with the Berkeley Drosophila Genome Project (BDGP).

melanogaster genome using these BACs. For further information

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

ECORI digestion of *Drosophila* DNA provided by the BDGP from the

P1 and EST libraries. A more detailed description of the library

filters for hybridization from the BACPAC Resource Center can be found at <http://bpac.org/bpac/bpac.html>.

FEATURES	
COURSE	Location/Qualifiers
1	025

```
/organism="Drosophila melanogaster"  
/mol_type="Genomic DNA"
```

```
/db_xref="taxon:7227"
/collection="BACB1916"
```

```
/clone_11b="RPCT-98"
/note="and : TET3"
```

ORIGIN

was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: M13Rev
LIBRARY

Vector : pKS145
R.site 1 : SacI
R.site 2 : SacI
Location/Qualifiers

1..1088
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-068F04.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 12.0%; Score 72.4; DB 10; Length 1088;
Best Local Similarity 46.0%; Pred. No. 0.00022;
Matches 220; Conservative 0; Mismatches 258; Indels 0; Gaps 0;

QY 1 GCCTCTACGCGCGGCAACGCTTCTTGAAGCCTTGGCCGTCGTGTCGACGACG 60
DB 803 GCG 744
QY 61 GGGCTGCGCGCTGACCTTGAAGCCTTGGCGGTCTGTGCGCGCGCGCGCGCGCGCG 120
DB 743 GGGCG 684
QY 121 GAGGCG 180
DB 683 GCG 624
QY 181 GAGGAGCTGCG 240
DB 623 GCGGCG 564
QY 241 CCGGAGCGGTGCTGCAACTGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 563 GCGGCG 504
QY 301 GGTGGGCGGTGCTGCG 360
DB 503 GCG 444
QY 361 GCGTGCATGCG 420
DB 443 GCG 384
QY 421 GCGGCGGTGCTGCG 478
DB 383 CCGGCG 326

RESULT 4

CNS006XK 935 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066051.1 GI:4945019

KEYWORDS

SSS. Drosophila melanogaster (fruit fly)

SOURCE

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 935)
Genoscope.
AUTHORS Direct Submission
TITLE

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeegawa and Aaron Mammeser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..935
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 11.9%; Score 71.8; DB 10; Length 935;
Best Local Similarity 29.4%; Pred. No. 0.00029;
Matches 122; Conservative 106; Mismatches 187; Indels 0; Gaps 0;

QY 114 CGGTACGAGGCG 173
DB 509 CGSCMCRYSGSGTCT 568
QY 174 GCGGATCGAGGAGCG 223
DB 569 CGGCTGCG 628
QY 234 CCGTACGAGGCG 293
DB 629 CGGCG 688
QY 294 CGAATCGGTGCG 353
DB 689 GCGGCG 748
QY 354 GCGGCGGTGCG 413
DB 749 SSSGCG 808
QY 414 CGAGGTGCG 473
DB 809 CGGCG 868
QY 474 CCGTACCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528
DB 869 GCGGCG 923

RESULT 5

CG756677/c 1033 bp DNA linear GSS 24-OCT-2003
LOCUS P051-4-D09.2b Ppa EcoRI BAC Library Pristionchus pacificus genomic.
DEFINITION genomic survey sequence.

ACCESSION

CG756677

KEYWORDS

SSS. Pristionchus pacificus

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasteridae; Neodiplogasteridae; Pristionchus.

REFERENCE

1 (bases 1 to 1033)
Srinivasan,J., Sins,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode *Pristionchus*
pacificus

JOURNAL
Mol. Genet. Genomics 269 (5), 715-722 (2003)

COMMENT

12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends
Location/Qualifiers
1. .1033
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Epa Ecoli BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with Ecoli and cloning into the BAC
vector."

FEATURES

source

Query Match 11.7%; Score 70.2; DB 10; Length 1033;
Best Local Similarity 45.4%; Pred. No. 0.0006;
Matches 249; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

ORIGIN

10 GCGCGCGGCAAGCCTTCTCTGAGCGCTTGGCCCGTGTGGTGGCGGAGTGGGCTGGCG 69
922 GCGAGGCG 863
70 GTCACTTGATGCTGCTGCGGCTCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
862 CGGCG 803
130 GACTACTTGCGCAGCG 189
802 GCG 743
190 CGAGCG 249
742 CGGCG 683
250 TTCTGCGAATGTTTCAACCG 309
682 CGGCG 623
310 CG 369
622 CGGCG 563
370 CGGAGCG 429
552 GCG 503
430 CTGGGCG 489
502 CGGCG 443
490 GACTCCATGACCG 549
442 CCG 383
550 GCCACGACC 558
382 GCGCGCGCGCG 374

RESULT 6

CNS0091P

LOCUS

CNS0091P 925 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #

DEFINITION

BACR19D16 of RPT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
AL053013 GI:4934461

KEYWORDS

AL053013.1 GI:4934461

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyrdoidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 925)

AUTHORS

Genoscope.

TITLE

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

JOURNAL

- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Ooegawa and
Aaron Mammoxer in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp; the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPT-98"
/note="end : TET3"

ORIGIN

Query Match 11.6%; Score 70; DB 10; Length 925;
Best Local Similarity 14.2%; Pred. No. 0.00066;
Matches 52; Conservative 175; Mismatches 139; Indels 0; Gaps 0;

233 ACTGGAACCGGACCGGTTCTGCACTGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292
524 ANNNNNAGCWSMCKKCGSTTBSTTTTTSAGSGVKGCGSGSGSCSCSCSCSCSCSCSCS 583
293 ACGAATCGGTGGGTCG 352
584 SCSCCBCCCCSCSYCCSSBSBSKCSSTBSBSCSCSSSVGTSVCSSSSSSS 643
353 GCG 412
644 TSSSTSSSTSSSSSSSSSSSSSYTTKSTBSAGSSWSAGSGSSTGSSSSSSSS 703
413 CCGAGTGGCAGCGGTGCTGGCGCAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 472
704 TSTSSSSVSBSKSTBSBSBSBSBSBSBSSTBSBBSCTSTSSSSSSSVSTSCCTC 763
473 TCCGTGACCTGGAGTTTCATTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 532
764 CCSYSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 823
533 TGACCGGCGGTCCGGGTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 592
824 TSTSCGSSSSSKGKGTTCGCGCGCGSSSTGMBGTSTACSSSSSSSSSVSSKSSA 883
593 CCGCGC 598

FEATURES
source
Location/Qualifiers
1..1007
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6270566"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 100"
/note="Organ: Liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 11.4%; Score 68.8; DB 5; Length 1007;
Best Local Similarity 43.7%; Pred. No. 0.0011;
Matches 248; Conservative 0; Mismatches 314; Indels 6; Gaps 1;
QY 27 CCTGACGACCTTCCGCGCGTGTGTCGCGACGAGTGGCTGCGGCTCACTTCGATGCGCTG 86
DB 426 CCGGGGCCCCCGGCG 485
QY 87 GGGTCTGTGGGCGCGGCGAAGCATGCGCGGTACCGAGGCGCGCGCATTAACCTGCGACCA 146
DB 486 GCG 545
QY 147 GGGCGTGGCGGCGCATGAGCG 206
DB 546 GGGCGTGGCGGCGCATGAGCG 605
QY 207 CCGGAGACCTTGGGCGTGTGCGTGTGAGACTGACCGGAGCGGTTCTGCAACTGTTTAC 266
DB 606 GGGGGGCG 665
QY 267 CCG 326
DB 666 GCG 725
QY 327 GACCGGTGAGAGATCGATCTCG 386
DB 726 GGGCG 785
QY 387 CGAGCATGTGCG 446
DB 786 C-----GCCCG 839
QY 447 GACGCGTGTGAGCGTGTGCG 506
DB 840 NCGGAGCG 899
QY 507 CGACCTGCGGAGACCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
DB 900 CCGCGGCG 959
QY 567 CGACCAACCGGAGCG 594
DB 960 GGGGCG 987

RESULT 11
CNS010EM/c 1009 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN03P19 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098882
VERSION AL098882.1 GI:5610493

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITILE
JOURNAL
COMMENT
GSS.
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

source

ORIGIN

Query Match 11.4%; Score 68.8; DB 10; Length 1009;
Best Local Similarity 33.5%; Pred. No. 0.0011;
Matches 123; Conservative 77; Mismatches 167; Indels 0; Gaps 0;
QY 46 CGTGTGCGGCGAGTGGGCTGCGGTCGATCGATCGGCTGTGAGCGCGGACG 105
DB 965 CCGCGMSRARGSAVTCGCGGCGCGGCGGMSAGCRSCCGGAGGSGSCSCSCSCSCS 906
QY 106 AACATGCGCGGTACCGAGGCGCGGCGCATCTTCGCGAGCGGAGCGCTGCGCGCATGAGAC 165
DB 905 CCSSGAGGCGSSGNGMSGSKGSRSMGCGGCGGSGGAGCGGCGSCGCCNCGRCGCG 846
QY 166 CCGGACCGGCGCATCGAGAGAGCTGCGGACCACTTGAAGCGCGGAGACCGCTGTGCTCG 225
DB 845 GCGACCGSSG 786
QY 226 GTGTGACCTGAGACCGGAGCGGCTTGTGTGAACTGTTCACCGCGCGCGCGCGCGCGCG 285
DB 785 CGCNSGSGSGGAGSGCGSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 726
QY 286 CTCTTGAAGAACTCGTGTGAGGTCGCGCGCGGCGCGGAGAGACCGGTACGATCGAT 345
DB 725 CCCG 666
QY 346 CTGCGCGCGCGCGTGTGCGGATGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 405
DB 665 SGGGSG 606
QY 406 GTCCGAG 412
DB 605 TTYBGKG 599

RESULT 12
AZ196424 886 bp DNA linear GSS 31-AUG-2000
LOCUS SP_1032.AL.A05.T7A Strongylocentrotus purpuratus, purple sea
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=1032 Col=9 Row=A, genomic survey sequence.
ACCESSION AZ196424
VERSION AZ196424.1 GI:8390247
KEYWORDS GSS.

[illegible]

OY	533	TGACCGGGATCCGGTGC	551
Db	839	GCGCGCGCGGGGCCCGC	857
RESULT 13			
LOCUS	CW949347	908 bp	DNA linear GSS 20-DEC-2004
DEFINITION	TcB33_4_G12_T7 Tribolium BAC library Tribolium castaneum genomic,		
ACCESSION	CW949347		
VERSION	CW949347.1		
KEYWORDS	GSS.		
SOURCE	Tribolium castaneum (red flour beetle)		
ORGANISM	Bukariyola, Metazoa; Arthropoda; Hexapoda; Insecta; Preyryota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium. 1 (bases 1 to 908) Savard,J. and Tautz,D. Tribolium castaneum BAC-ends sequencing project unpublished (2003) Contact: Savard, J. Abteilung fur Evolutionsgenetik, AG Tautz Institut fur Genetik, Universitat zu Koln Meyercal 121, 50931 Koln, Germany Tel.: 49 221 470 6911 Fax: 49 221 470 5975 Email: savard@uni-koeln.de Class: BAC ends.		
FEATURES			
source	1..908	Location/Qualifiers	
	/organism="Tribolium castaneum"		
	/mol_type="genomic DNA"		
	/strain="GA-2"		
	/db_xref="taxon:7070"		
	/clone_id="Tribolium BAC library"		
	/note="vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;		
	Library constructed by Exelixis Inc."		
ORIGIN			
Query Match	11.0%; Score 66.4; DB 10; Length 908;		
Best Local Similarity	47.5%; Pred. No. 0.0034;		
Matches	209; Conservative 0; Mismatches 228; Indels 3; Gaps 1		
OY	34	GCTTCGCGCCGTGCGTGCAGTGAGGCGCGCACCTGCATGCGCTGGGGTCTG	93
Db	415	GCGCGCGGCGCGCCCGCGCGCGCTGCGGAGGCGGGGCGCCCCCGCGCGCG	474
OY	94	TGGGCGGCGAGAACATGAGCGGTATCCAGAGGCGCGCATACTGCGAGCCAGCGGCGCTG	153
Db	475	CNCGGGGGGGCGCGCGCGGCGGGGGCGGGGGGGCGCGCGCGCGCGCGGGGGGGGG	534
OY	154	CGCGGCATGAGCCCGCAGCGGGCGATTCAAGAGCTGCGGACCACTTGGAGCGCCGGGGAC	213
Db	535	CGGGCGGGGGGCGCCNCGCGCGCGCGGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG	594
OY	214	CGGTGGGTGTGGTGGTGGAACTTGACCGGGAGCGGTTGATGAATGTGTACAACCGCGCC	273
Db	595	GCGGCGGGGGGCGGGGCGGGGGGGGGGGGGCGGGGCGCGCGCGCGCGCGCGCGCGCG	654
OY	274	CGCGCCCGGCGCTCTTC--GACAACTCGTGGGGTCCGCGCGGGGCGAGGAGACC	330
Db	655	CGCGCGCGGGGGCGCGCGCGGGGGGCGCGGGGGGGGGGGGGCGCGCGCGCGCGCGCG	714
OY	331	GGTCAAGAATCGAATCTGCGCCCGCGCGCTGGGCTGGATGCGCGAGAGCGGAAGTACAG	390
Db	715	GGGCGCGCGCGCGCGCGCGCGCGCGCGGGGGGGGGGGGGCGGGGCGCGGGGCGGGG	774
OY	391	CATGTCGCGCGCGCTGTGTCGAGCCGAGGTGCGAGCGGTGCTGGGCGACGCGCGGACG	450
Db	775	NCGGCGCGCGCGCGCGCGCGGGGGGGGGGGGGCGCGCGCGCGCGCGCGGGGGGGCGCG	834

451 GTGATCGAGCGTGACGTCC 470
 AG371264
 AG371264
 854 GGGCGCGCGGCGGCGCGGCG 854

RESULT 14
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AG371264 1547 bp DNA linear GSS 21-DEC-2004
 Mus musculus molossinus DNA, clone:MSMg01-176J21.TU, genomic survey
 sequence.
 AG371264
 AG371264.1 GI:47982469
 GSS.
 Mus musculus molossinus (Japanese wild mouse)
 Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Abe,K., Noguchi,H., Tagawa,K., Yuzurika,M., Toyoda,A., Kojima,T.,
 Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
 Shiotoishi,T.
 Contribution of Asian mouse subspecies Mus musculus molossinus to
 genomic constitution of strain C57BL/6J, as defined by BAC-end
 sequence-SNP analysis
 Genome Res. 14 (12), 2439-2447 (2004)
 15574823
 2 (bases 1 to 1547)
 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 Direct Submission
 Submitted (17-NOV-2003) Maahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-Chou,Tsukumi-Ku, Yokohama, Kanagawa, 230-0045, Japan
 (E-mail:hattori@sc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the mouse BAC library MSMg01. For BAC
 library availability, please contact Kunita Aoe (aoe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: aoe@rtc.riken.jp

COMMENT
 PRIMERs
 Sequencing : TU
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
 Location/Qualifiers
 1..1547
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-176J21.TU"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

FEATURES
 SOURCE

11.0%; Score 66; DB 10; Length 1547;
 Best Local Similarity 47.7%; Pred.No. 0.004;
 Matches 246; Conservative 0; Mismatches 266; Indels 4; Gaps 2;

31 GAGCGCTTGCGCCGCGTGTGCGCGCGAGTGGCTGCGCGTCACTGATGCGCTTGGGT 90
 |||||
 637 GCGCGCGGCGGCGGCC 696
 |||||
 91 CTGTGTGCGCGCGGAGACATGCGCGCGTATCCAGAGGCGCGAGCTACTGTGCGAGCGACG 150
 |||||
 697 GCGGCGGCGGCG 756
 |||||
 151 CTGGCGCGCATGAGACCGCGAGCGGCGCATGAGAGAGCTGCGGACACCGCTTGAACCGCGG 210

[illegible]

Db	1380	GAGGAGCTGGCGGACCACTCTGGACGCGCGGGGACCCCTGTGGTGTGGTGTGGACCTTGGAC	1439
Qy	241	CGGAGCGGGTTCTGTGAACCTGTTCAACCGCGCGCCCGCGCGGCGCCCTTTCGAGGAATCTC	300
Db	1440	CGGAGCGGGTTCTGTGAACCTGTTCAACCGCGCGCCCGCGCGGCGCCCTTTCGAGGAATCTC	1499
Qy	301	GGTGGGGTCCGCGCGCGGGGCGGAGGAGACCGGATCGAATCGAATCTTCGCGCGGCGCTG	360
Db	1500	GGTGGGGTCCGCGCGCGGGGCGGAGGAGACCGGATCGAATCTTCGCGCGGCGCTG	1559
Qy	361	GGCTTCGAATGCGGAGGCGGAAACGTCAAGAGCATGTCCGCGCGGCTGTCCGAGCGGAGGTG	420
Db	1560	GGCTTCGAATGCGGAGGCGGAAACGTCAAGAGCATGTCCGCGCGGCTGTCCGAGCGGAGGTG	1619
Qy	421	GGAGCGGGCTGGGGCCAGGGACCGCGACCGGATCGAGCGTGAAGCTTCGCTTCGTTGAC	480
Db	1620	GGAGCGGGCTGGGGCCAGGGACCGCGACCGGATCGAGCGTGAAGCTTCGCTTCGTTGAC	1679
Qy	481	CTGGGATTCGACTCCATGACCGCGCGTGCAGCTTCGCGGAAACCGGCTTCGCGGCGGTGACCGGG	540
Db	1680	CTGGGATTCGACTCCATGACCGCGCGTGCAGCTTCGCGGAAACCGGCTTCGCGGCGGTGACCGGG	1739
Qy	541	GTCCGGGTGGCCACGACCATGTGTCTTCGACCAACCCGACAGTGGACCGGCTTCACCGCGCAC	600
Db	1740	GTCCGGGTGGCCACGACCATGTGTCTTCGACCAACCCGACAGTGGACCGGCTTCACCGCGCAC	1799
Qy	601	T 601	
Db	1800	T 1800	

```

RESULT 2
US-10-672-396-5
/ Sequence 5, Application US/10672396
/ Publication NO. US20040166567A1
/ GENERAL INFORMATION:
/ APPLICANT: SANITI, DANIEL V.
/ APPLICANT: REID, RALPH C.
/ APPLICANT: KODUMAL, SARAH J.
/ APPLICANT: JAYARAJ, SRBASTIAN
/ TITLE OF INVENTION: SYNTHETIC GENES
/ FILE REFERENCE: 3006220109300
/ CURRENT APPLICATION NUMBER: US/10/672,396
/ CURRENT FILING DATE: 2003-09-26
/ PRIOR APPLICATION NUMBER: US 60/14,085
/ PRIOR FILING DATE: 2002-09-26
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: FastSeq for Windows Version 4.0.
/ SEQ ID NO 5
/ LENGTH: 9510
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic construct
US-10-672-396-5

```

Query Match	46.8%	Score 281	DB 7	Length 9510
Best Local Similarity	66.7%	Pred. NO. 5e-60		
Matches 401; Conservative	0	Mismatches 200	Indels 0	Gaps 0

QY	1	GGCTCTTACGGCGGGGCAAGCGCTTCTTCAGCGCTTGGCCGCTGTGTGTGGGGCACT	60
Db	3778	GGCTCTTACGGCGGGGCAAGCGCTTCTTCAGCGCTTGGCCGCTGTGTGTGGGGCACT	3833
QY	61	GGGCTGCGGCTAAGCTGATGGCTGTGGGTCTGTGGGCGGGGAGAAATATGGCGGTACC	120
Db	3838	GGCGACCGCTACAGATATCCATGGGGGTGTGGGCGGCTCAGAACTGGCGGTGAT	3893
QY	121	GAGGGCGCGGATACCTTGCGGACGAGGGCGCGCGCATGGAATCCGAGGGGCGATC	180
Db	3898	GAGGCGGGATATCTCGGTAGCCAGGGCTGGCGCGCATGAAATCCAGATGTGGGGG	3953

OY	181	GAGGAGCTGGCGGACACACCTTGGAACGCGGGGACCCGGTGGGTGGGTGGGACTGGAC	240
Db	3958	GAAAGACTGCATATCAAGCTGGATCAAGTTCAGACTTCGGTCTAAGTGTGATATGGAC	4017
OY	241	CGGAGCGGTTTCGTGAACTGTTACAGCGCGCCGCGCGGCGCCCTTTCGAGAACTC	300
Db	4018	CGTGGCGGTTTTGTGGAGTTGTTTCAACGAGCTGCCGCTACACGCGCTTTGTGATGAATC	4077
OY	301	GATGGGGTCCGCGCGCGGGGCGGAGGAACCGGTGAAGAAATCGAATCTCCGCGCGGCTG	360
Db	4078	GCGGATGACGAGGAGGAAGCTCCGCAAGTGAAGAGGAGCGCTGCGCTGGCGCAAGCTTG	4137
OY	361	GCGTCGATGCCGAGGAGCGGAACTGCACAGAGCATGTCCGCCGCTGATCCGAGCGAGTG	420
Db	4138	GCOCGACTGTCTACCGGCGAGCGCGCGGAGGACCTGGGACACTGATTCGGTCCGGAAGTG	4197
OY	421	GCAGCGGTGCTGGGCGCACGCGCACCGCGACGCTGATTCGAGCTGACGTCGCTTCCGTGAC	480
Db	4198	GCACCGGCTTTCGTGTCAACGCGACGATGCCGCGATTGACCCGCAATCGTGCATTTCCGCGAT	4257
OY	481	CTGGGATTCGACTGCATGACCGCGCGTGGACCTTGCGGGAACCGGCTCGCGGCGGTGACCGGG	540
Db	4258	CTGGGATTGACTCAATGACTGCGGTGACCTTGCGGCAACGCTCTTCGCAAGCGGTGACCGGG	4317
OY	541	GTCGCGGTGGCACGAGCAACCATGTCCTTGACCAACCCGACAGTGAACCGCTTCAACGCGCAC	600
Db	4318	GTAAGTAGGCTGCGCAACGTTGTTATTGACCATCAACGATCAACGCGCTTGGCGGATCAT	4377
OY	601	T 601	
Db	4378	T 4378	

```

RESULT 3
US-10-205-032-17
; Sequence 17, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-17

```

Query Match	36.6%	Score 220.2	DB 5	Length 4725
Best Local Similarity	61.6%	Pred. No. 5e-45		
Matches 370, Conservative	0	Mismatches 228	Indels 3	Gaps 1

QY	3	CTCTACACGGGGGGGCAACGGCTTCTCTGACGGCTTCCGCGTCTGCTGCGCGCATGTCG	62
Db	3879	CGCGTACGGGGGGGCGCAACGGCGGCGCTCTGACCGCGTCCCGCAACCGCGCGCGCGCGCGCG	3938
QY	63	GCTGCGCGGTCACTTCGATCGCTCGTGGGTCTGTGGCGCGGCGCAACATGCGCGGTACCGA	122
Db	3939	GCTGACCGCGCACTTCGTCCGCTCGCGGGGCTGTGGGGGCTTCGCGGGGCGATGGCGGAGGCGA	3998
QY	123	GGGCGGGCACTACCTGCGCGACGCAAGGCGCTTGCGCGCATGACCCGCGAGCGGGCGAATCGA	182
Db	3999	CGCCGAGAGTACTTGACCGCGCGGGGCGCTGCGCGCATGTGCTCCGACGCGTGGCGTGGGA	4058
QY	183	GGAGCTGGGGACCAACCTCGTGAACGCGCGGGGACCCGTTGGGTGTCGATGAGTGGACCTTGACCG	242
Db	4059	CGCCCTCTTGGCGCGGCTTGACCGGGACGAGACTTCCTCGTCCGCTGCGCGCATGTGGAATTG	4118

Query Match	Best Local Similarity	36.6%;	Score 220.2;	DB 5;	Length 60196;
Matches 370;	Conservative	0;	Mismatches 228;	Indels 3;	Gaps 1;
Qy	3	CTCTTACGGGAGCGGGACAAGCGCTTCTCTCGAAGCGCTTGCGCCGTGTGTGTGCGGCGAGTGG	62		
Db	40170	CGCCTTACGGGGCGGSCCAACGGGGCCCTTCGACGGCTTCGAAACGGGCGGCGCCCGCG	40225		
Qy	63	GCTGCGCGATCACCTCGATCGCTTGTGGGTGTGTGGGCGGGACAACATGCGCGGTACCGA	122		
Db	40230	GCTAACCGGCACCTCCGTCGCTGTGGGCGCTGTGGGCTCTCGGGGGGCAATGGSCAAGGCGA	40285		
Qy	123	GGGGCGGCGACTTCTTGGCAGCGAGGGCGCTTGCAGCATGAAACCGCAGCGGCGCGATTCGA	182		
Db	40290	CGCGAGAGAGTAACTGTAGCGCGCGGGGCTCTGGGGCGGATGCTCTCCGAGGTGGCGGTGA	40345		
Qy	183	GGAGCTGCGGACCAACCTTGAACGCGGGGGAACCGGTGGTGTGAGTGTGAACCTGGAACCG	242		
Db	40350	CGCCCTTCTGGCGCGCCCTTGAACCGGAGCGAGACCTTGTGTGCGCGTGCAGACTGTGACTG	40405		
Qy	243	GGAGCGGTGTGTGGAATGTTCACCGCGCGCGCGCGCGCGCCCTCTTTCGACGAATTCGG	302		
Db	40410	GACCTGTGTTCACGCGCGGATTCACCGGTTCGAGCCGACGAGCGGCTGTCTGTGGTCAACTCC	40465		

Query Match	34.9%	Score 210	DB 6	Length 18438
Best Local Similarity	61.3%	Prod. No. 1	2e-42	
Matches 356	Conservative 0	Mismatches 220	Indels 6	Gaps 1

QY	1	GCCCTCTACCGGGGCGCAACGCTTCTCTGACGCTTGCCCGTCGTGAGTCCGCGCACT	60
Db	11284	GGGCGCGTACCGCGCGCCGCAACGGTATCGCCGACGCCCTCGCGGACCGCGCGCAAGCGCC	11341
QY	61	GGGCTGCGCGTCACTCGATCGCTTGAGGGTCTGTGTGGCGCGGCAGAACATGCGCCGTATCC	120
Db	11344	GGCTTCACCGCCACCTTCGCGGCTCTGGGGGACGTGGGCGCGCGCGGATGGTTCAGCGAC	11403
QY	121	GAGGGCGGCGACATACTGTGGCGAGCGGACGTGGCGGCATATGACCCCGAGGCGCGGATC	180
Db	11404	GACCTGGCGCGGAACTTCGCGCCCGCGCGGGGTGCGCTTCATATGACCCCGAGCGCGGATC	11463
QY	181	GAGGAGCTGCGGACCAACCTTGAGCGCCGCGGACCCGTGGGTGTCCGTGTGTGACCTTGAC	240

Db	1164	GGCGCGCTCCAGCAGGCGCTTCGACCAACGACGAGACCAACGCTGACGGTTCACCGACATGAGC	11523
Qy	241	CGGAGAGCGGTTTCGTTCGAATCTGTTACACGCGCGCCCGCGCGCGCTCTTTCGACGAATTC	300
Db	11524	TGGAGAGCGCTTCGCGCGGAGAGTTTCACCGCGCGCCCGCGCGCTCTTCATTCGACGCAATC	11583
Qy	301	GGTGGGGTTCGCGCGCGCGGGCC-----GAGGAGACCGGTTCAGGAATCGAATCTCGCGCGG	354
Db	11584	CCGAGAGCGCGCGCGCGCGCGCAACGAGAAATCCGCGCGCGGAGAAATTCGAATCTGCGCGCC	11643
Qy	355	CGGCTGCGCTGATGATGCGCGGAGCGCAACGTTCACGAGATGTCGCGCGGCTGTGTCGAGGC	414
Db	11644	AGGCTGCGCGGGTTCGACCGACGCGGAAACGGAAACGGAATCTCTTCGACTCTGTGACGAGCC	11703
Qy	415	GAGGTGGCAAGCGATGCTGGGCGCAACGACACCGACGATATCGAGCGTGAACGTCCCTTC	474
Db	11704	GCGCGCGCGCTCGCGCTGGGGCACCGGGGCGCGAGGCCATTCACCCCGTGAAGCCCTTC	11763
Qy	475	CGTGAACCTGGGATTCGACTCGATTCATGACCGCGCTGCACTTCGGGAAACCGGCTTCGGGGGTG	534
Db	11764	AAGGACCTGGGCTTCGACTCCCTGACCGCGCTGCACTTCGGGCAACCGGCTTCACCGCGCG	11823
Qy	535	ACCGGGGTCGGGGTGGCGACGACCATGTCCTTCGACCAACCGGAC	578
Db	11824	ACCGGCGTACTCTCCCGCGACCGCTGTCTTGGACCAACCCGAC	11867

```

RESULT 6
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

```

Query Match	34.9%;	Score 210;	DB 6;	Length 125746;
Best Local Similarity	61.3%;	Pred. No. 8.4e-43;		
Matches	358;	Conservative	0;	Mismatches 220; Indels 6; Gaps 1;

QY	1	GCTCTTACGCGCGCGGCAACGCGCTTCTCTGACGCGCTTGCCTGCTGCTGCGCGCAGT	60
DB	36594	GCGCGGTACGCGCGCGCGCAACGCGCTTACGCGACCGCGCTCCGCGCGGACCGCGCGCAAGCGCGC	36553
QY	61	GGGCGTCCCGGTCACCTCGATCGCTGCGGGGTCTGTGTGGCGCGGGCAGAAATATGCGCGGTACC	120
DB	36654	GGCTTCATCGCCACTCTCTGTTGAGCTTGGGGCAGCTTGGGCGCGCGCGGAGATGTCTGACGAC	36713
QY	121	GAGGCGGCGCACTACCTTGCAGCGCAGAGGCGCTTGTGCGCCATATGACCCGACGCGGGCGATC	180
DB	36714	GACCTGGCGCGGAGACTTGCACCGCGCGCGGGGTGGCTCGATGACCCGACCGGCGGATC	36772
QY	181	GAGGAGCTCGCGGACCACTCTGGAAGCGCGCGGACCCCTGTGGGTCTCGTGTGTGACCTTGGAC	240
DB	36774	GCGCGCTTCACAGCAGGCGCTTGCACCGACGACGCTGACGGTGCATCCGACCATATGAGC	36833

OY	241	CGGAGGCGTTCGTGCACTGTTCAACCGCCGCCCGCGCGCCCTCTTGACGAACTC	300
Db	36834	TGGAGAGGCTTCCCGCGAGCGTTCACTCGGCGCCCGCGCCCTCTTACGACGCATC	36893
OY	301	CGTGGGGTCCGCGCCCGGGGCC-----GAGAGACCGGTCAGGAATCGATCTCGCCCG	354
Db	36894	CCGAGGCGCGCCCGCGCCCGCCACGACGAGATCCGGCCGCGGAGATCCGAATCGGCGCGC	36955
OY	355	CGGCTGCGCTGCATGCGCGGAGCGCAAGTCAAGACATGTGCGCCCGGCTGGTCCGAGCC	414
Db	36954	AGGCTGCGCGGGGTGCACGACGGCGAAACGGGAACGGGAATCTCGTCACTGGTCAGAGACC	37013
OY	415	GAGGTGCGACGGTGTCTGGGCGACGGGCACGGCGACGGTATCGAGGCTGACGTGCTTC	474
Db	37014	GCCGCGCCCTTCGCGCTGGGGGACCGGGGCGCGGAGCGCATCACCCCGTGGAGCCCTTTC	37073
OY	475	CGTGAACCTTGGGAATTCCATCTGCATGACCGCGCTGCGACGTGCGGAAACCGGCTCGCGGCGGTC	534
Db	37074	AAGGACCTTGGGCTTTCGACTCTCTGACCGCGCTGACCTTGGGCACACGGGCTCACCGCGCGC	37133
OY	535	ACCGGGTCCGGGTGCGACGACCATGTGTTTGACCAACCGGAC	578
Db	37134	ACCGGCTTACGTCTGCCCGCGCAAGCTGTGTTTGACCAACCGCAC	37177

```

1 RESULT 7
2 US-10-156-761-1/c
3 ; Sequence 1, Application US/10156761
4 ; Publication No. US20030119018A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: OMOURA, SATOSHI
7 ; APPLICANT: IKEDA, HARUO
8 ; APPLICANT: ISHITAKA, JUN
9 ; APPLICANT: HORIKAWA, HIROSHI
10 ; APPLICANT: SHIBA, TADAYOSHI
11 ; APPLICANT: SAKAKI, YOSHIYUKI
12 ; APPLICANT: HATTORI, MASHIIRA
13 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
14 ; FILE REFERENCE: 249-262
15 ; CURRENT APPLICATION NUMBER: US/10/156, 761
16 ; CURRENT FILING DATE: 2002-05-29
17 ; PRIOR APPLICATION NUMBER: JP 2001-204089
18 ; PRIOR FILING DATE: 2001-05-30
19 ; PRIOR APPLICATION NUMBER: JP 2001-272697
20 ; PRIOR FILING DATE: 2001-08-02
21 ; NUMBER OF SEQ ID NOS: 15109
22 ; SEQ ID NO 1
23 ; LENGTH: 9025608
24 ; TYPE: DNA
25 ; ORGANISM: Streptomyces avermitilis
26 ; FEATURE:
27 ; NAME/KEY: misc feature
28 ; LOCATION: (4187715)
29 ; OTHER INFORMATION: a, t, c, g, other or unknown
30 US-10-156-761-1

```

	Query Match	Similarity	Score	DB	Length
Beet Local	34.9%	61.3%	210	3.6e-43	9025608
Matches	358	Conservative	0	Mismatches	220
				Indels	6
				Gaps	1

QY 181 GAGAGCTGCGGACCACTTGAACCGGAGAACCGGAGTGTCTGTGTGAACCTGAC 240
Db 3614885 GCGCGCTCCAGCAGGCGCTTCCAGCAAGAGACAGCGCTGACCTGCAACGAGAC 3614826
QY 241 CCGGAGCGGTTCTGTGAACCTTTCACCGCGCGCGCGCGCGCGCGCTTCTTTCAGCAATC 300
Db 3614825 TGGAGAGGCTTCCGCGGAGAGTTTCAACCGCGCGCGCGCGCGCGCGCTTCTTTCAGCAATC 3614766
QY 301 GGTGGAGTCCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
Db 3614765 CCGGAGCG 3614706
QY 355 CCGCTGCGCTGATGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
Db 3614705 AGGCTGCG 3614646
QY 415 GAGTGGCGACCGGTTCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474
Db 3614645 GCG 3614586
QY 475 CGTGAAGTGGAGTTGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
Db 3614585 AAGGACCTGCGCGCTTCCAGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614526
QY 535 ACCGCGGTCGCGGTCG 578
Db 3614525 ACCGCGGTCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614482

RESULT 8
US-10-760-493-34
Sequence 34, Application US/10760493
Publication No. US20050187167A1

GENERAL INFORMATION:
APPLICANT: Ecopia Biosciences Inc
APPLICANT: Farnet, Chris M.
APPLICANT: McAlpine, James B.
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Bachmann, Brian O.
APPLICANT: Pirae, Mahmood
TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
FILE REFERENCE: 3004-905
CURRENT APPLICATION NUMBER: US/10760,493
CURRENT FILING DATE: 2004-01-21
PRIOR APPLICATION NUMBER: USSN 60/441,123
PRIOR FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: USSN 60/469,810
PRIOR FILING DATE: 2003-05-13
PRIOR APPLICATION NUMBER: USSN 60/491,516
PRIOR FILING DATE: 2003-08-01
PRIOR APPLICATION NUMBER: USSN 60/494,568
PRIOR FILING DATE: 2003-08-13
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 9684
TYPE: DNA
ORGANISM: Streptomyces albusensis
US-10-760-493-34

Query Match 34.5%; Score 207.6; DB 9; Length 9684;
Best Local Similarity 60.3%; Pred. No. 5,4e-42;
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

QY 1 GCGCTCTACGCGCGGAGCAAGCGCTTCTGACGCGCTTCCGCGCGCGCGCGCGCGCGCGCG 60
Db 4054 GCGAATACGCGCGCGGAGCAAGCGCTTCTGACGCGCTTCCGCGCGCGCGCGCGCGCGCG 4113
QY 61 GCGCTCTACGCGCGGAGCAAGCGCTTCTGACGCGCTTCCGCGCGCGCGCGCGCGCGCGCG 120
Db 4114 GCGAATACGCGCGCGGAGCAAGCGCTTCTGACGCGCTTCCGCGCGCGCGCGCGCGCGCG 4173

QY 121 GAGGCGCGGACTTACCTGCG 180
Db 4174 GCGCTGTGATGCGGATGCG 4233
QY 181 GAGAGCTGCGGACCACTTGAACCGGAGAACCGGAGTGTCTGTGTGAACCTGAC 240
Db 4234 GCGCACTCAGCAGCGCGCTTCCAGCAAGAGACAGCGCTGACCTGACCGTGTGCGCAATC 4293
QY 241 CCGGAGCGGTTCTGTGAACCTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 4294 TGGAGAGGCTTCCGCGGAGAGTTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4353
QY 301 GGTGGAGTCCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336
Db 4354 CCGGAGCG 4413
QY 337 GAATGAGTCTGCG 396
Db 4414 GCGCGCGCGCTGCG 4473
QY 397 GCGCGCGCTGCTGCG 456
Db 4474 CTGGAATGCTGCTGCG 4533
QY 457 GAGCGTGAAGTGGCTTCCG 516
Db 4534 GAGCG 4593
QY 517 AACCGCTGCG 576
Db 4594 AACCGCTGAGCG 4653
QY 577 ACAATGACCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 4654 ACCCGCACCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4677

RESULT 9
US-10-760-493-18
Sequence 18, Application US/10760493
Publication No. US20050187167A1

GENERAL INFORMATION:
APPLICANT: Ecopia Biosciences Inc
APPLICANT: Farnet, Chris M.
APPLICANT: McAlpine, James B.
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Bachmann, Brian O.
APPLICANT: Pirae, Mahmood
TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
FILE REFERENCE: 3004-905
CURRENT APPLICATION NUMBER: US/10760,493
CURRENT FILING DATE: 2004-01-21
PRIOR APPLICATION NUMBER: USSN 60/441,123
PRIOR FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: USSN 60/469,810
PRIOR FILING DATE: 2003-05-13
PRIOR APPLICATION NUMBER: USSN 60/491,516
PRIOR FILING DATE: 2003-08-01
PRIOR APPLICATION NUMBER: USSN 60/494,568
PRIOR FILING DATE: 2003-08-13
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.0
SEQ ID NO 18
LENGTH: 164051
TYPE: DNA
ORGANISM: Streptomyces albusensis
US-10-760-493-18

Query Match 34.5%; Score 207.6; DB 9; Length 164051;
Best Local Similarity 60.3%; Pred. No. 3.1e-42;
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

Db 62752 GAGGGGAAACGCTGCTGCGCCCGGCTGCTGTAAGGCGATGAGCCGCGCGGCGCTG 62811
Qy 181 GAGAGGTGCGAACAACCTTGAACGCGGGGAAACCCCTGGGTGTCTGGTGTGACCTGAC 240
Db 62812 GCGCACTGGGCGCGGCGCTGTGTGAAGACCTTCACTGTGAAGGCGCGACCTTGAC 62871
Qy 241 CGGAGCGGTTCGTGCACTGTTCAACGCGCGCGCGCGCGCGCGCTTCTTCAAGAACT- 239
Db 62872 CGGCGCGATTCGCGGCGGCGCTACACTCGCGCGCTTCCGCGCGCTGATCGCGAACTG 62931
Qy 300 -----CGTGGGGTCCGCGCGCGGCGCGAGAGAACCGGTCAGAATCTTCCGCGG 354
Db 62932 ATGACGCGGAGCGCGCGACCGCGACCGCGCGCGCGCGCGCGCGGAGTGTGGAC 62991
Qy 355 CGGCTGGCGTCAATGCGCGAGGCGGAACGTCAGAGATGTCCGCGCGCTGTGCGAGCC 414
Db 62992 CGGCGGTGACCGCGCTGCGCGCGCGCGCTGCGCGCGAACTGTCTCACTGTGTCCGCGC 63051
Qy 415 GAGTGGCAAGGCTGTGCGCGCAACGCGACCGCGAGTGTGAGCGTGTGCGCTTC 474
Db 63052 GAGTGGCGCGCAACTCGCGCAACGCGGCGGTGAGGCGCATCGAACCGACCGGCGCTTC 63111
Qy 475 CGTACCTGGGATTCGATCTCATGACCGCGCTGCGAGAACCGGCTGCGCGCGGTG 534
Db 63112 CGGACCTCGGCTTCACTGCTGTGCGCGCGCGCTGCGAGCTGCGAACCGGATTCGCGAGGCC 63171
Qy 535 ACCGGGTGCGGGTGGCGCAACGACCATGTCTTGAACCAACCGACGATGAGCCGCTCAC 594
Db 63172 ACCGGGTGACACTGTGCGCGCGCGCGCGCTGATCTACGACCAAGAACCGCGCGCTGCGC 63231
Qy 595 GCGCAC 600
Db 63232 GCGACAC 63237

RESULT 12
US-10-760-493-22
; Sequence 22, Application US/10760493
; Publication No. US20050187167A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc
; APPLICANT: Farnet, Chris M.
; APPLICANT: McAlpine, James B.
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Bachmann, Brian O.
; APPLICANT: Pirae, Mahmood
; TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE
; FILE REFERENCE: 3004-9US
; CURRENT APPLICATION NUMBER: US/10/760,493
; CURRENT FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: USN 60/441,123
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: USN 60/469,810
; PRIOR FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: USN 60/491,516
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: USN 60/494,568
; PRIOR FILING DATE: 2003-08-13
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 24444
; TYPE: DNA
; ORGANISM: Streptomyces albusensis
US-10-760-493-22

Query Match 33.1%; Score 199; DB 9; Length 24444;
Best Local Similarity 60.5%; Pred. No. 5,9e-40;
Matches 378; Conservative 0; Mismatches 220; Indels 27; Gaps 2;
Qy 3 CTCCTACGGGCGGCAAGCCTTCTCTCAAGCCTTCCGCGCGTGTGTGCGGAGTGG 62

Db 22559 CGCTTACGCGGCGGCAATGCGCACTTGAGCGCGCTTGCGCCACAGCGCGCGCGCG 23618
Qy 63 GCTCGCGGTCACTCGATGCGCTTGGGCTGTGGGCGGCGAGAACTG--GCCGTAC 119
Db 23619 ACTGAGCGGCGCTCGGTGCGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 23678
Qy 120 CGAGGCGGCGGCTTACCTGTGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
Db 23679 CGAGGCGGCGGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 23738
Qy 180 CGAGGAGCTGCGAGCAACCTTGAGCGCGGAGAACCGGTGATGCGGTGAGCTTGA 239
Db 23739 CGCGCGCGCTTCAAGAGCGGTGAGCGGAGCAAGAGCGGTGACGCTGCGCATGTGCA 23798
Qy 240 CGGAGACCGGTTCGTGAACTGTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
Db 23799 CTGGGACCTGTGTGTGCTCGGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCG 23858
Qy 300 CGGTGGGGTCCG-----GCCGGGCGGAGAGCCGCTCA 335
Db 23859 CCGCGAGGTGCGCGCGCTCTGTGCGGCAAGAGAGAGCGCGCGCGCGCGCGCGGGA 23918
Qy 336 GGAATGGAATCTTGCCTGCGCGCGCGCTGCGATGCGGAGGCGGAACGTCAAGAGATGT 395
Db 23919 AGCGGCGCGCTTGAAGCGGAGCTGCGGGGATGAGCGGAACCGAGCGGAGGCGTGT 23978
Qy 396 CGCGCGCTGTGTGCGAGCGGAGGTGCAAGCGGTGCTGCGGCGCAAGCGCGCGCGTAT 455
Db 23979 CTGGAACCTGTGTGCTGAGCAAGGTGCGCGGTCTTCTGAGCAAGGGGAGCGAGCGGT 24038
Qy 456 CGAGGCTGACGTGTGCTTCCGTGACCTTGGGATTCATCATGACCGCGGTGACCTTGG 515
Db 24039 CGAGGCGCGCGCGCTTGAAGGACCTGCGCTTCACTGCTCAACCGCGGTGACGTGCG 24098
Qy 516 GAACGCGCTGCGGCGGAGTACCGGCGGTGCGGAGCGCAAGACCATGTCTTGAACACCC 575
Db 24099 CAACGCGCTGACGACCGCGCAACCGGACTGCGGCGTCCGCGAGCGCTGTCTTGAACCA 24158
Qy 576 GACAGTGAACCGCTTCAACCGCGCAC 600
Db 24159 GACCGCGCGCGCACTGCGCGCGCAC 24183

RESULT 13
US-10-156-761-412
; Sequence 412, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 412
; LENGTH: 18435
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(18435)
US-10-156-761-412

Query Match 33.0%; Score 198.4; DB 6; Length 18435;
Best Local Similarity 60.4%; Pred. No. 8.8e-40;
Matches 378; Conservative 0; Mismatches 221; Indels 27; Gaps 2;

```
Qy 1 GCGCTCTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCGCGAGT 60
Db 13027 GCGGCTTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCGGCGGCGCA 13086
Qy 61 GGGCTGCGGCTGACCTGATGCTGCGGCTGCTGCGGCTGCGGCGGCGGCGGCGGCGG 120
Db 13087 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13146
Qy 121 GAGGCGGCGGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177
Db 13147 GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13206
Qy 178 ATCGAGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
Db 13207 GTGCGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13266
Qy 238 GACCGGCGGCGGCTTCTGCGAACTGTTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db 13267 GACTGCGGCGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 13326
Qy 298 CTGCGTGGGCGG-----TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333
Db 13327 CTCCCGGAGGCGGCGGAGTCTTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13386
Qy 334 CAGGAATCGATCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393
Db 13387 GCGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13446
Qy 394 GTCGCGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453
Db 13447 CTGCTGATCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13506
Qy 454 ATCGAGGCTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513
Db 13507 GTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13566
Qy 514 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
Db 13567 CGGAACGAGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13626
Qy 574 CGGACGCTGACCGCGCTTCAACCGCGCA 599
Db 13627 CCGAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13652
```

RESULT 14
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000

TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Query Match 33.0%; Score 198.4; DB 6; Length 100000;
Best Local Similarity 60.4%; Pred. No. 6.3e-40;
Matches 378; Conservative 0; Mismatches 221; Indels 27; Gaps 2;

```
Qy 1 GCGCTCTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCGCGAGT 60
Db 63550 GCGGCTTACGCGGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCGGCGGCGGCGGCGG 63609
Qy 61 GGGCTGCGGCTGACCTGATGCTGCGGCTGCTGCGGCTGCGGCGGCGGCGGCGGCGG 120
Db 63610 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63669
Qy 121 GAGGCGGCGGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177
Db 63670 GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63729
Qy 178 ATCGAGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
Db 63730 GTGCGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63789
Qy 238 GACCGGCGGCGGCTTCTGCGAACTGTTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db 63790 GACTGCGGCGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 63849
Qy 298 CTGCGTGGGCGG-----TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333
Db 63850 CTCCCGGAGGCGGCGGAGTCTTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63909
Qy 334 CAGGAATCGATCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393
Db 63910 GCGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63969
Qy 394 GTCGCGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453
Db 63970 CTGCTGATCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64029
Qy 454 ATCGAGGCTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513
Db 64030 GTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64089
Qy 514 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
Db 64090 CGGAACGAGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64149
Qy 574 CGGACGCTGACCGCGCTTCAACCGCGCA 599
Db 64150 CCGAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64175
```

RESULT 15
US-10-156-761-411
; Sequence 411, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

This Page Blank (uspto)

```

RESULT 1
US-09-679-279-1
Sequence 1, Application US/09679279
Patent No. 6524841
GENERAL INFORMATION:
APPLICANT: McDaniel, Robert
APPLICANT: Volchegursky, Yana
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
TITLE OF INVENTION: Genes and Uses Thereof
FILE REFERENCE: 300622004700
CURRENT APPLICATION NUMBER: US/09/679,279
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158,305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190,024
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicina
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (978)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosiaminyl transferase (eryCIII homolog),
OTHER INFORMATION: TDP-megsamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase;
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDII, deoxybysugar transaminase (eryCI, DnrJ homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminoctransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS

```

LOCATION: (5822)...(6595)
OTHER INFORMATION: megd11, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megd1v, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmuh hc
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megdV, TDP-hexose 4-ketoreductase (eryBIV, dmuv homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megB11-1(megdVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megA1, SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megA1, AT-L
NAME/KEY: misc feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megA1, KS1
NAME/KEY: misc feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megA1, AT1
NAME/KEY: misc feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megA1, KR1
NAME/KEY: misc feature
LOCATION: (17917)...(18207)
OTHER INFORMATION: megA1, ACP1
NAME/KEY: misc feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megA1, KS2
NAME/KEY: misc feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megA1, AT2
NAME/KEY: misc feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megA1, KR2
NAME/KEY: misc feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megA1, ACP2
NAME/KEY: CDS
LOCATION: (22867)...(33555)
OTHER INFORMATION: megA11, SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (32957)...(4237)
OTHER INFORMATION: megA11, KS3
NAME/KEY: misc feature
LOCATION: (42544)...(25581)
OTHER INFORMATION: megA11, AT3
NAME/KEY: misc feature
LOCATION: (26230)...(26733)
OTHER INFORMATION: megA11, KR3 (inactive)
NAME/KEY: misc feature
LOCATION: (26958)...(27258)

OTHER INFORMATION: megA11, ACP3
NAME/KEY: misc feature
LOCATION: (27353)...(28590)
OTHER INFORMATION: megA11, KS4
NAME/KEY: misc feature
LOCATION: (28897)...(29931)
OTHER INFORMATION: megA11, AT4
NAME/KEY: misc feature
LOCATION: (29953)...(30477)
OTHER INFORMATION: megA11, DH4
NAME/KEY: misc feature
LOCATION: (31396)...(32244)
OTHER INFORMATION: megA11, ER4
NAME/KEY: misc feature
LOCATION: (32257)...(32799)
OTHER INFORMATION: megA11, KR4
NAME/KEY: misc feature
LOCATION: (33052)...(33312)
OTHER INFORMATION: megA11, ACP4
NAME/KEY: CDS
LOCATION: (33666)...(43271)
OTHER INFORMATION: megA11, SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (33780)...(35027)
OTHER INFORMATION: megA11, KS5
NAME/KEY: misc feature
LOCATION: (35385)...(36419)
OTHER INFORMATION: megA11, AT5
NAME/KEY: misc feature
LOCATION: (37068)...(37604)
OTHER INFORMATION: megA11, KR5
NAME/KEY: misc feature
LOCATION: (37860)...(38120)
OTHER INFORMATION: megA11, ACP5
NAME/KEY: misc feature
LOCATION: (38187)...(39470)
OTHER INFORMATION: megA11, KS6
NAME/KEY: misc feature
LOCATION: (39755)...(40811)
OTHER INFORMATION: megA11, AT6
NAME/KEY: misc feature
LOCATION: (41406)...(41936)
OTHER INFORMATION: megA11, KR6
NAME/KEY: misc feature
LOCATION: (42168)...(42425)
OTHER INFORMATION: megA11, ACP6
NAME/KEY: misc feature
LOCATION: (42585)...(43271)
OTHER INFORMATION: megA11, TE
NAME/KEY: CDS
LOCATION: (43268)...(44344)
OTHER INFORMATION: megA11, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (44355)...(45623)
OTHER INFORMATION: megA11, desosaminyl transferase, desosamine glycosyltransferase
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (45620)...(46591)
OTHER INFORMATION: megB11-2(megB11), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 2,3 dehydratase;
OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (46660)...(47403)
OTHER INFORMATION: megH, TE11, SEQ ID NO: 19= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (47411)...(47980)
OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ
us-09-679-279-1

Query Match 100.0%; Score 601, DB 3, Length 47981;
Best Local Similarity 100.0%; Pred. No. 2,5e-104, Indels 0, Gaps 0;
Matches 601, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY	1	GCCTCTACGCGGCGGGAACGCTTCTCTGAGCGCTTGGCCGCTGCTGGCGCACT	60
Db	37506	GCCTCTACGCGGCGGGAACGCTTCTCTGAGCGCTTGGCCGCTGCTGGCGCACT	37506
QY	61	GCGCTGCCGAGTCACCTTCGATCGCTTGGGAGTCGTGTGGCCGGGAGAACATGCGCGTACC	120
Db	37566	GCGCTGCCGAGTCACCTTCGATCGCTTGGGAGTCGTGTGGCCGGGAGAACATGCGCGTACC	37566
QY	121	GAGGCGCGCGCATCTACCTGCGGAGCGCAGGAGCGCTGCGCCATGGAACCCGACGCGGCGATC	180
Db	37626	GAGGCGCGCGCATCTACCTGCGGAGCGCAGGAGCGCTGCGCCATGGAACCCGACGCGGCGATC	37626
QY	181	GAGGAGCTGCGGAGCAACCTTCGAGCGCGGAGGACCCGCGGAGTGTGCGGTGTGAGCTTGAC	240
Db	37686	GAGGAGCTGCGGAGCAACCTTCGAGCGCGGAGGACCCGCGGAGTGTGCGGTGTGAGCTTGAC	37686
QY	241	CGGAGCGCGGTTGCTGCAACTGTTCACCGCGCGCCCGCGCGCCCTCTTCGACGAACTC	300
Db	37746	CGGAGCGCGGTTGCTGCAACTGTTCACCGCGCGCCCGCGCGCCCTCTTCGACGAACTC	37746
QY	301	GCTGGGCTTCCG	360
Db	37806	GCTGGGCTTCCG	37806
QY	361	GCGTCGATGCGCGAGCGCGGACGTCGACGATGTCGCGCGCGCGCGCGCGCGCGCGCGCG	420
Db	37866	GCGTCGATGCGCGAGCGCGGACGTCGACGATGTCGCGCGCGCGCGCGCGCGCGCGCGCG	37866
QY	421	GCACCGGCTGCTGGGCGACCGCGCACCGCGCACCGGTCGAGCGTCGCTTCGTCGAC	480
Db	37926	GCACCGGCTGCTGGGCGACCGCGCACCGCGCACCGGTCGAGCGTCGCTTCGTCGAC	37926
QY	481	CTGGGATTTCCATCTTCATGACCGCGCTTCGACCTTCGAGAACCGGCTCGCGCGGTCGAC	540
Db	37986	CTGGGATTTCCATCTTCATGACCGCGCTTCGACCTTCGAGAACCGGCTCGCGCGGTCGAC	37986
QY	541	GTCGCGGTGCGCACGACCATCTGCTTCGACCAACCGGACGTCGACCGGCTTCGCGCGAC	600
Db	38046	GTCGCGGTGCGCACGACCATCTGCTTCGACCAACCGGACGTCGACCGGCTTCGCGCGAC	38046
QY	601	T 601	
Db	38106	T 38106	

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckerts, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..33921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl-ACP synthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydroatase and enoylreductase domains m"
FEATURE:

```

NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of acyl carrier domain of module 4"
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: beta-ketocylACP synthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of acyltransferase domain of module 5"
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of acyl carrier domain of module 5"
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of beta-ketocylACP synthase domain of modul"
OTHER INFORMATION: beta-ketocylACP synthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of acyltransferase domain of module 6"
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19145..19398
OTHER INFORMATION: /function= "approximate span of acyl carrier domain of module 6"
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of thioesterase domain of module 6"
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match 66.2%; Score 397.8; DB 2; Length 20235;
Best Local Similarity 78.9%; Pired. No. 3.3e-66;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Db 1 GCCTCTACGCGGCGGCAAGCCTTCTCTGAGCCTTGGCCGCTGTGTGCGGCGAGT 60
14500 GCTCTCTACGCGGCGGCAAGCCTTCTCTGAGCCTTGGCCGCTGTGTGCGGCGAGT 14559
Qy 61 GGGCTGCGGCTCACTCTGATGCGCTGTGGGTCTGTGGGCGGCGAGAACATGAGCGGTACC 120
14560 GGGCGCGCGGTGACGTTCATGCTGGGGGCTGTGGGCGGCGAGAACATGAGCGGAGAC 14619
Qy 121 GAGGGCGGCGACTACTCTGCGACACGAGGCTTGGCGGCTTGAACCGGCGGCGATC 180
14620 GAGGGCGGCGAGTACTCTGCGACACGAGGCTTGGCGGCTTGAACCGGCGGCGATC 14679
Qy 181 GAGAGGTGCGGACCACTCTGACGCGCGGAGACCGGTGGGTCTGGGTGTGGAACCTTGAC 240
14680 GAGAGTGTGACATCACTCTGACCACTGACCACTGACCACTGACCACTGACCACTGAC 14739
Qy 241 CGGAGCGGCTTCTGCTGCACTGTTTCACTGCGCGCGCGCGCGCGCTTCTTGAAGATC 300
14740 CGGAGCGGCTTCTGCTGCACTGTTTCACTGCGCGCGCGCGCGCGCTTCTTGAAGATC 14799
Qy 301 GGTGGGGTCCG 360
14800 GCGGCTCCGCTC 14859
Qy 361 GCGTCATATGCGGAGCG 420
14860 GCGGCTGTCGACG 14919
Qy 421 GCAAGCGGTGCTGGGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
14920 GCGCGGCTGCTGCG 14979
Qy 481 CTGGGATTGCACTTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
14980 CTGGGCTTGCATCTTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15039
Qy 541 GTCCGGGTGCG 600
15040 GTCCGGGAGCG 15099
Qy 601 T 601
Db 15100 T 15100
RESULT 3
US-08-439-009A-3
Sequence 3, Application US/08439009A
Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSES: Steven F. Weinstein
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/439,009A
3 FILING DATE: 11-MAY-1995
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Casuto, Dianne
7 REGISTRATION NUMBER: 40,943
8 REFERENCE/DOCKET NUMBER: 4952.US.D1
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: 847-938-3137
11 INFORMATION FOR SEQ ID NO: 3:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 20235 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: double
16 TOPOLOGY: unknown
17 MOLECULE TYPE: DNA (genomic)
18 HYPOTHEICAL: NO
19 ANTI-SENSE: NO
20 ORIGINAL SOURCE:
21 ORGANISM: Saccharopolyspora erythraea
22 STRAIN: NRRL 238
23
24 FEATURE:
25 NAME/KEY: CDS
26 LOCATION: 19..10722
27 OTHER INFORMATION: /codon_start= 19
28 OTHER INFORMATION: /function= "gene eryA"
29 OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
30 OTHER INFORMATION: 6-deoxyerythronolide B"
31 FEATURE:
32 NAME/KEY: misc feature
33 LOCATION: 19..4470
34 OTHER INFORMATION: /function= "approximate span of
35 OTHER INFORMATION: module 3"
36 FEATURE:
37 NAME/KEY: misc feature
38 LOCATION: 97..1482
39 OTHER INFORMATION: /function= "approximate span of
40 OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
41 FEATURE:
42 NAME/KEY: misc feature
43 LOCATION: 1693..2670
44 OTHER INFORMATION: /function= "approximate span of
45 OTHER INFORMATION: acyltransferase domain module 3"
46 FEATURE:
47 NAME/KEY: misc feature
48 LOCATION: 3406..3921
49 OTHER INFORMATION: /function= "approximate span of
50 OTHER INFORMATION: beta-ketoreductase domain of module 3"
51 FEATURE:
52 NAME/KEY: misc feature
53 LOCATION: 4171..4428
54 OTHER INFORMATION: /function= "approximate span of
55 OTHER INFORMATION: acyl carrier domain of module 3"
56 FEATURE:
57 NAME/KEY: misc feature
58 LOCATION: 4471..10722
59 OTHER INFORMATION: /function= "approximate span of
60 OTHER INFORMATION: module 4"
61 FEATURE:
62 NAME/KEY: misc feature
63 LOCATION: 4471..5847
64 OTHER INFORMATION: /function= "approximate span of
65 OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
66 FEATURE:
67 NAME/KEY: misc feature
68 LOCATION: 6054..7026
69 OTHER INFORMATION: /function= "approximate span of
70 OTHER INFORMATION: acyltransferase domain of module 4"
71 FEATURE:
72 NAME/KEY: misc feature
73 LOCATION: 7165..9216
74 OTHER INFORMATION: /function= "approximate span of

1 OTHER INFORMATION: dehydratase and enoylreductase domains m"
2 FEATURE:
3 NAME/KEY: misc feature
4 LOCATION: 9433..9984
5 OTHER INFORMATION: /function= "approximate span
6 OTHER INFORMATION: beta-ketoreductase of module 4"
7 FEATURE:
8 NAME/KEY: misc feature
9 LOCATION: 10225..10483
10 OTHER INFORMATION: /function= "approximate span of
11 OTHER INFORMATION: acyl carrier domain of module 4"
12 FEATURE:
13 NAME/KEY: CDS
14 LOCATION: 10723..20235
15 OTHER INFORMATION: /codon_start= 10723
16 OTHER INFORMATION: /function= "gene =eryA"
17 OTHER INFORMATION: /product= "orf13 encoding modules 5 & 6
18 OTHER INFORMATION: 6-deoxyerythronolide B formatio"
19 FEATURE:
20 NAME/KEY: misc feature
21 LOCATION: 10723..15165
22 OTHER INFORMATION: /function= "approximate span of
23 OTHER INFORMATION: module 5"
24 FEATURE:
25 NAME/KEY: misc feature
26 LOCATION: 10831..12174
27 OTHER INFORMATION: /function= "approximate span of
28 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
29 FEATURE:
30 NAME/KEY: misc feature
31 LOCATION: 12375..13350
32 OTHER INFORMATION: /function= "approximatr span of
33 OTHER INFORMATION: acyltransferase domain of module 5"
34 FEATURE:
35 NAME/KEY: misc feature
36 LOCATION: 14062..14610
37 OTHER INFORMATION: /function= "approximate span of
38 OTHER INFORMATION: beta-ketoreductase of module 5"
39 FEATURE:
40 NAME/KEY: misc feature
41 LOCATION: 14857..15114
42 OTHER INFORMATION: /function= "approximate span of
43 OTHER INFORMATION: acyl carrier domain of module 5"
44 FEATURE:
45 NAME/KEY: misc feature
46 LOCATION: 15166..20235
47 OTHER INFORMATION: /function= "approximate span of
48 OTHER INFORMATION: module 6"
49 FEATURE:
50 NAME/KEY: misc feature
51 LOCATION: 15172..16569
52 OTHER INFORMATION: /function= "approximate span of
53 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
54 FEATURE:
55 NAME/KEY: misc feature
56 LOCATION: 16768..17721
57 OTHER INFORMATION: /function= "approximate span of
58 OTHER INFORMATION: acyltransferase domain of module 6"
59 FEATURE:
60 NAME/KEY: misc feature
61 LOCATION: 18375..18921
62 OTHER INFORMATION: /function= "approximate span of
63 OTHER INFORMATION: beta-ketoreductase domain of module 6"
64 FEATURE:
65 NAME/KEY: misc feature
66 LOCATION: 19145..19398
67 OTHER INFORMATION: /function= "approximate span of
68 OTHER INFORMATION: acyl carrier domain of module 6"
69 FEATURE:
70 NAME/KEY: misc feature
71 LOCATION: 19492..20235
72 OTHER INFORMATION: /function= "approximate span of
73 OTHER INFORMATION: thioesterase domain of module 6"

US-08-439-009A-3

Query Match 66.2% Score 397.8; DB 3; Length 20235;
Best Local Similarity 78.9%; Pred. No. 3.3e-66;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
QY 1 GCGTCTACGCGGCGGCAACGCTTCTCGACGCTTCGCCGCTCGTGTGCGCGAGT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14500 GCGTCTACGCGGCGGCAACGCTTCTCGACGCTTCGCCGCTCGTGTGCGCGAGT 14559
QY 61 GGGCTCGCGTCACTTGATGCTGTGGGTCTGTGGGCGCGGCGAACAATGGCGGTATC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14560 GGGCTCGCGTCACTTGATGCTGTGGGTCTGTGGGCGCGGCGAACAATGGCGGTATC 14619
QY 121 GAGGGCGGCGCACTACCTGCGGACGAGGGCTTGCGCGCAATGACCCGCGGCGGATC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14620 GAGGGCGGCGCACTACCTGCGGACGAGGGCTTGCGCGCAATGACCCGCGGCGGATC 14679
QY 181 GAGGAGCTGCGGACCACTTGAAGCGCGGAGACCGGTGCTGTGCTGTGATCTTGAC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14680 GAGGAGCTGCGGACCACTTGAAGCGCGGAGACCGGTGCTGTGCTGTGATCTTGAC 14739
QY 241 CGGAGCGGCTTGTGCACTGTTCACCGCGCGCGCGCGCGCGCGCTTCTTGACAAATC 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14740 CGGAGCGGCTTGTGCACTGTTCACCGCGCGCGCGCGCGCGCGCTTCTTGACAAATC 14799
QY 301 GGTGGGGTCCGCGCGCGCGCGCGGAGAGACGGGTGAGAAATCGAATCTCGCGCGGCTG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14800 GGTGGGGTCCGCGCGCGCGCGCGGAGAGACGGGTGAGAAATCGAATCTCGCGCGGCTG 14859
QY 361 GCGTGTGATGCGGAGCGGCAAGTTCACGAGCATGTGCGCGCGCTGTGCGAGCGAGGTG 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14860 GCGTGTGATGCGGAGCGGCAAGTTCACGAGCATGTGCGCGCGCTGTGCGAGCGAGGTG 14919
QY 421 GCAAGCGGTGCTGGGCGACGCGCACCGCGCATGATGAGCGGTGACGTGCTTCCGTGAC 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14920 GCGCGGTGCTGGGCGACGCGCACCGCGCATGATGAGCGGTGACGTGCTTCCGTGAC 14979
QY 481 CTGGGATTCGACTCATGACCGCGGTGACCTGCGGAAACGGGCTCGCGCGGTGACCGCGG 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14980 CTGGGATTCGACTCATGACCGCGGTGACCTGCGGAAACGGGCTCGCGCGGTGACCGCGG 15039
QY 541 GTCCGCGGTGCGACGACCATGCTTTCGACCAACCGGACGAGTGTGACCGCGCGGAC 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15040 GTCCGCGGTGCGACGACCATGCTTTCGACCAACCGGACGAGTGTGACCGCGCGGAC 15099
QY 601 T 601
DB 15100 T 15100
```

RESULT 4

US-07-642-734C-1

Sequence 1, Application US/07642734C

Patent No. 5824513

GENERAL INFORMATION:

APPLICANT: Katz, L

APPLICANT: Donadio, S

APPLICANT: McAlpine, J B

TITLE OF INVENTION: Recombinant DNA Method for Producing

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSER: Edward H. Gorman

STREET: Abbott Laboratories D377/AP6D-2 One Abbott

CITY: Abbott Park

STATE: IL

COUNTRY: US

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/642,734C

FILING DATE: 17-JAN-91

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Danckers, Andreas M

REGISTRATION NUMBER: 32652

REFERENCE/DOCKET NUMBER: 4952.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396

TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 11219 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Saccharopolyspora erythraea

STRAIN: NRRL 2338

FEATURE:

NAME/KEY: misc feature

LOCATION: 744..6659

OTHER INFORMATION: /function= "APPROXIMATE SPAN OF

OTHER INFORMATION: MODULE 1"

OTHER INFORMATION: /label= FUNCTION

FEATURE:

NAME/KEY: CDS

LOCATION: 744..11219

OTHER INFORMATION: /function= "gene= "eryA"

OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for

OTHER INFORMATION: 6-deoxyerythronolide B"

FEATURE:

NAME/KEY: misc feature

LOCATION: 744..1868

OTHER INFORMATION: /function= "approximate span of

OTHER INFORMATION: acyltransferase domain 1 of module 1"

OTHER INFORMATION: acyl carrier domain 1 of module 1"

FEATURE:

NAME/KEY: misc feature

LOCATION: 3831..4811

OTHER INFORMATION: /function= "approximate span of

OTHER INFORMATION: acyltransferase 2 domain of module 1"

OTHER INFORMATION: beta-ketoadacylACP synthase domain/module1"

FEATURE:

NAME/KEY: misc feature

LOCATION: 6369..6626

OTHER INFORMATION: /function= "approximate span of

OTHER INFORMATION: acyl carrier domain 2 of module 1"

OTHER INFORMATION: module 2"

FEATURE:


```

? NAME/KEY: misc feature
? LOCATION: 6678..8066
? OTHER INFORMATION: /function= "approximate span of
? OTHER INFORMATION: beta-ketoacyl AcP synthase of module 2"
?
? FEATURES:
? NAME/KEY: misc feature
? LOCATION: 8262..9305
? OTHER INFORMATION: /function= "approximate span of
? OTHER INFORMATION: acyltransferase domain of module 2"
?
? FEATURES:
? NAME/KEY: misc feature
? LOCATION: 9906..10454
? OTHER INFORMATION: /function= "approximate span of
? OTHER INFORMATION: beta-ketoreductase domain of module 2"
?
? FEATURES:
? NAME/KEY: misc feature
? LOCATION: 10707..10964
? OTHER INFORMATION: /function= "approximate span of
? OTHER INFORMATION: acyl carrier domain of module 2"
?
US-07-642-734C-1

```

Query Match	33.7%;	Score 202.6;	DB 2;	Length 11219;
Best Local Similarity	60.9%;	Pred. No. 1.4e-29;		
Matches 366;	Conservative	0;	Mismatches 229;	Indels 6; Gaps 2

QY	1	GCTCTTACAGCGGCGGGGCAACGCTTCTCTGAGAGCCTTGGCCGCTGTGGTGGGCGCACT	60
Db	10344	GGCCCTTACGCGCGGGGCAACGCTTCTCTGAGAGCCTTGGCCGCTGTGGTGGGCGCACT	10403
QY	61	GGGCTGCGCGATTCACCTTCATTCGCTTGGGATCTGTGGCCGGGAGAGAACATATGGCCGTACC	120
Db	10404	GGCTGCGCGGCAAGTGGATGGCGTGGGGGCTGTGGCGGGCGGGCGGCAATATACCGGCGAAC	10463
QY	121	GAGGGCGCGCG---ACTTACCTTGGGAGGCAAGGCGCTGTGGCGCCATATGAAACCGGACGGGGG	177
Db	10464	GAGGAGGCGCGGTGCTGTCTTTCGCGGAGCGGGGTGGGGGAGATGCGCCGTATACCGCGCGCC	10523
QY	178	ATTCAGAGAGCTGCGGAGCAACCTTCAGACGCGCGGGAGACCGGTGGGTGTTCGGTGTGAACTTG	237
Db	10524	CTGCGCCGCTTCAGACAGGGGTGCTGGCTTCGGGGAGACGGCGGGTGTCTGTACAGACAGCTTG	10583
QY	238	GACCGGAGAGCGGTTTCGTGCACCTGTTCACCGCGCGCCGCGCGGCGCCCTTCYTTCAGAA	297
Db	10584	GACTGGCGCGGCTTTCGCGGAGTCTTACACGCGCGCCGCGCGCGGTGTTCAGACCGCG	10643
QY	298	CTCGGTGGGGTCCGCGCGGGGCGGAGAGAACCGGTTCAGGAATTCGGATCTTCGCGCG---G	354
Db	10644	ATTCCTCACAGACCGCGCGGAGCGAGCGGGCGGAGAACCGGAGACGAGAGGCTTCGCGCGAAC	10703
QY	355	CGGCTGCGCTGCATGCGCGGAGGCGGAACTTCAGAGACATATTCGCGCGGCTGTGTCGAGGCG	414
Db	10704	CGGCTGCGCGGTTCTGCGCGGTGCGGAGCGGAGCGGAGAGCTGTGTGGCTGTGCTGCACAC	10763
QY	415	GAGGTGGCAGCGCGTGTCTGGGCGACAGGCAACGCCGACCGGTATTCAGCGGTATCGTGCCTTTC	474
Db	10764	AGCAACCGCGACCGGTCTGGGCGCACAGACCTCCAGAGCGGTGGCGCGGACCAACGCGGTTTC	10823
QY	475	CGTACCTTGGAGATTTCAGCTCCATGACGCGCGGTGAGACTGTGGAGAACGGGCTCCGCGCGGTGG	534
Db	10824	AMGAGCTTCGGGTTTCAGCTGCTGGGCGGCGCGGTGCGGCAACTGTGCTCAACGGGGGCGC	10883
QY	535	ACCGGGGTTCGGGTGGCGACAGACATATGTTTCAGACCAACCGGACAGATGAGACGGCTCACCC	594
Db	10884	ACCGGGGTTCGGGTCGTGCGAGCGGTGTTCAGACCAACCGGAGCGCTCCGCGGTGCGC	10943
QY	595	G	595
Db	10944	G	10944

RESULT 5
US-08-439-009A-1
; Sequence 1, Application US/08439009A

```

Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSER: Steven F. Weinstein
STREET: Abbott Laboratories D377/AP6d-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Caetano, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc feature
LOCATION: 744..6659
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc feature
LOCATION: 744..1668
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 3831..4811

```


Dh 39481 GGTCTCCGCGCACTCGATCTCTGAGGAGTTGGGCGGCGGCGCATGCGGACGGC 39540
Qy 121 GAGGCGGCGGACTTACCTGCGCAGCAGGCGCTGCGGCCATGGAACCGCGGAGGATC 180
Dh 39541 GCGGCGGCGGAACTTGGCGGCGAAGCGGATGATGCGATGCGGCGGCGGCGCATC 39600
Qy 181 GAGAGCTGCGGACCACTTGAAGCGCGGAGACCGGTGGTGTGCGTGTGAGACTGAC 240
Dh 39601 CTGGCTCTGAGAGATCTTGAACAAGATGAGACGTGGGTGTGATGCGTATGTGAC 39660
Qy 241 CCGGAGCGGCTCTGCACTGTTCAACCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 300
Dh 39661 TGGAGCCGATTCTTCCACGTTTCGCGCATCTGCGGCAACCGGTTGTTCAGAGATG 39720
Qy 301 GGTGGGATCCGCGCGCGGCGGCGGAGAGACCGGTCAAGAA-----TCGATCTC 348
Dh 39721 CCGGCGGCGGAGAAAGCGGATGCGCGGATGAGCGCGGAGAACAGGCGGCTGCGGTT 39780
Qy 349 GCGCGGCGGCTGCGCTGATGCGGAGGCGGAACTGACAGAGATGCGCGCGCTGATC 408
Dh 39781 GCGCGGAACTCTGCGGAGGCTGCGGAGGCGGAAAGCGCAAGCGCACTGTGTGATCTG 39840
Qy 409 CGAGCGGAGGTGCGAGCGGCTGCGGCGCAAGCGGCGGCGGATGATGAGCGTGAAGTC 468
Dh 39841 TGGCGCCAGGTGCGAACCGGTCTCGGCGCAAGCGGATGCGGAGGATCCAGCGCGG 39900
Qy 469 GCGTTCGCGGACCTGAGATTCATCTCATGACCGCGGTGACCTTGGGAAACCGGCTCGG 528
Dh 39901 GCGTTCGCGGCGCTCGGATTCGATCTCTCATGCGGATGATGCGGATCTTGTGAC 39960
Qy 529 GCGGTGACCGGAGTCCGCGGTGCGCACAGACATGCTTGCAGCACCGGACAGTGAAGCG 588
Dh 39961 ACCGCGACCGGAGTGGCGCTGCGGACCAACACCGTCTTGACATCCGGAATCGGCGGC 40020
Qy 589 CTCACCGGCGAC 600
Dh 40021 TTGGCGGCTCAC 40032

RESULT 7

US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 32.5%; Score 195.6; DB 3; Length 80161;
Best Local Similarity 59.0%; Pred. No. 2.8e-28;
Matches 361; Conservative 0; Mismatches 239; Indels 12; Gaps 1;
Qy 1 GCCTCTACGCGGCGGCAAGCGCTTCTCGACGCTTGGCGCGGTGCGTGGCGGCACT 60
Dh 39421 GCGGCTTACGCGGCGGCGGCAAGCGATTTCTTGACACACTGCGGCAACAGCGGAGCAGCGC 39480

Qy 61 GGGCTCCGAGTCACTCGATGCGCTTGGGATCTGTGGGCGCGGCGAGAACATGCGCGATCC 120
Dh 39481 GGTCTCCGCGCACTCGATCTCTTGGGCACTTGGGCGGCGGCGGCGGCGGCGGCGG 39540
Qy 121 GAGGCGGCGGACTTACCTGCGCAGCAGGCGCTGCGGCGCAATGGAACCGCGGAGATC 180
Dh 39541 GCGGCGGCGGAACTTGGCGGCGAAGCGGATGATGCGATGCGGCGGCGGCGGCGCATC 39600
Qy 181 GAGAGCTGCGGACCACTTGAAGCGCGGAGACCGGTGGTGTGCGTGTGAGACTGAC 240
Dh 39601 CTGGCTCTGAGAGATCTTGAACAAGATGAGACGTGGGTGTGATGCGTATGTGAC 39660
Qy 241 CCGGAGCGGCTCTGCACTGTTCAACCGCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 300
Dh 39661 TGGAGCCGATTCTTCCACGTTTCGCGCATCTGCGGCAACCGGTTGTTCAGAGATG 39720
Qy 301 GGTGGGATCCGCGCGCGGCGGCGGAGAGACCGGTCAAGAA-----TCGATCTC 348
Dh 39721 CCGGCGGCGGAGAAAGCGGATGCGCGGATGAGCGCGGAGAACAGGCGGCTGCGGTT 39780
Qy 349 GCGCGGCGGCTGCGCTGATGCGGAGGCGGAACTGACAGAGATGCGCGCGCTGATC 408
Dh 39781 GCGCGGAACTCTGCGGAGGCTGCGGAGGCGGAAAGCGCAAGCGCACTGTGTGATCTG 39840
Qy 409 CGAGCGGAGGTGCGAGCGGCTGCGGCGCAAGCGGCGGCGGATGATGAGCGTGAAGTC 468
Dh 39841 TGGCGCCAGGTGCGAACCGGTCTCGGCGCAAGCGGATGCGGAGGATCCAGCGCGG 39900
Qy 469 GCGTTCGCGGACCTGAGATTCATCTCATGACCGCGGTGACCTTGGGAAACCGGCTCGG 528
Dh 39901 GCGTTCGCGGCGCTCGGATTCGATCTCTCATGCGGATGATGCGGATCTTGTGAC 39960
Qy 529 GCGGTGACCGGAGTCCGCGGTGCGCACAGACATGCTTGCAGCACCGGACAGTGAAGCG 588
Dh 39961 ACCGCGACCGGAGTGGCGCTGCGGACCAACACCGTCTTGACATCCGGAATCGGCGGC 40020
Qy 589 CTCACCGGCGAC 600
Dh 40021 TTGGCGGCTCAC 40032

RESULT 8

US-09-603-207-1
; Sequence 1, Application US/09603207B
; Patent No. 6521406
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/603,207B
; EARLIER FILING DATE: 2000-06-23
; EARLIER APPLICATION NUMBER: 09/370,700
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-603-207-1

Query Match 32.5%; Score 195.6; DB 3; Length 80161;
Best Local Similarity 59.0%; Pred. No. 2.8e-28;
Matches 361; Conservative 0; Mismatches 239; Indels 12; Gaps 1;
Qy 1 GCCTCTACGCGGCGGCAAGCGCTTCTCGACGCTTGGCGCGGTGCGTGGCGGCACT 60
Dh 39421 GCGGCTTACGCGGCGGCGGCAAGCGATTTCTTGACACACTGCGGCAACAGCGGAGCAGCGC 39480

Db 39421 GGGGCTTACGGCGCGGGAACGATCTTGGACACACTGCGCGAACAACGCCGAGCACGC 39480
QY 61 GGGGCTGCGGCTCACCTTGATCGCTTGGGCTCTGTGGCCCGGGGAGAACTAGGCGGATC 120
Db 39481 GGTCTGCGCGCAACCTCGATCTCTTGGGAGAGTGGGCGCGCGGCAATGGCGGACGCGC 39540
QY 121 GAGGGCGGCGCACTACCTGCGGAGCAGGGGCTGGCGGCATGGAACCCGCGAGCGGCGATC 180
Db 39541 GCGCGGCGCGCAACCTGCGCGGAGCGCGGAGATGCTCCATGCGCGGCGGCTGGCGCATC 39600
QY 181 GAGAGCTGCGGACCACTTGAACGCGCGGAGCCCGTGGTGTGAGTGAACCTTGAC 240
Db 39601 CTGGCTTGCGAGGAAGTACTTGAACCAAGATGAGACGTGCGTGAATCGCTGAATGAGAC 39660
QY 241 CGGAGCGGCTTCTGTCGAATGTTTCAACCGCGCGCGCGCGCGCGCGCTTCTTGAACCTC 300
Db 39661 TGGGACCGGATTCGTTCCACGTTTGGCGGCACTGCGCGCACCCGCTTGTTCGAAGAGTG 39720
QY 301 GGTGGGGTCCGCGCGCGGCGCGGAGAGACCGGTCAGGA-----TGGATCTC 348
Db 39721 CCGCGCGGAGAAAGCGATGCCCGCGATGGCGCGGAGAACCAAGCGCGCTCGCGCTTC 39780
QY 349 GCCCGCGCGCTGCGCGTGAATGCCGAGAGCCGAAAGCTCAAGACATGTGCGCGCGCTGCTC 408
Db 39781 GCCCGCAATCTCGCGGAGCTGCGCGAAGCCCAACGACGCACTGTGATCTGATG 39840
QY 409 CGAGCGGAGTGGCGAGCGGCTGCTGGCGGCGCGGACGCGGAGTGTGAGCGTGAACGTC 468
Db 39841 TGCAGCGGAGTGGCGAAGCGCTGCGGCGGAGTGGCGGAGAGTCCAGCGCGGAGCGG 39900
QY 469 GCGTTCGCGTGAAGTGGGATTCGATCCATGACCGCGCGTGAAGTGGGAGACCGGCTGCGG 528
Db 39901 GCGTTCGCGCGGCTGCGGTTGATCTCTTCAAGCGGAGTGTGCGGATCTGTTGAC 39960
QY 529 GCGGTGAACCGGAGTCCGCGTGGCGGACGACGATGCTTGAACCAACCGGAGAGTGAACGC 588
Db 39961 ACCCGCAACCGGAGTGGCGCTGCGCGGACGACGATGCTTGAACCAACCGGAGAGTGGCGG 40020
QY 589 CTCACCGCGCAC 600
Db 40021 TTGGCGCGCTCAC 40032

RESULT 9
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuester, Stuart A.
; APPLICANT: Kuester, Paul R., Jr.
; APPLICANT: Sulton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSER: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII (DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 30.3%; Score 182.4; DB 2; Length 4437;
Best Local Similarity 60.1%; Pred. No. 8.3e-26;
Matches 363; Conservative 0; Mismatches 226; Indels 15; Gaps 3;

QY 7 TACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCGCGTGTGTGCGCGGCGAGTGGCTG 66
Db 35220 TATCGCGCGCGCAATGCGCATCTGAGTGGCGTGGCGGAGCGTGTGCGCGAGGGCGT 35279
QY 67 CCGGTCACTTGATGCGCTTGGGCTGTGTGGCGCGGCGCAAACTATGCGCGTACCGAGGC 126
Db 35280 CCGCGACCTTCCGCTGCGCTGCGGCGCTGTGGGCGCGCGAGGCGCATGAGGAGGAGGC 35339
QY 127 GGC---GACTACCTGCGCGACGCGGCGCTGCGCGGCGCATGGAACCGCGGCGAGTGA 183
Db 35340 GTCAAGGAGTGTACGCGCGAGCGCGCTTGCCTCCCATGCGGCGCGAGTGGCGATGAG 35399
QY 184 GAGCTGCGGACCACTTGAACGCGCGGAGACCGGTGGTGTGTGAGTGAACCTGAGCGG 243
Db 35400 GCACTTGAACAGGCACTGGAAGAGGCGGACACTGCGTCAACGCTGCGCGCATGATG 35459
QY 244 GAGCGGTTCTGGAATGTTTCAACGCGCGCGCGCGCGCGCGCTTCTTGAACGAACTGGT 303
Db 35460 GAACACTTCTGCAACGGGTTTCAACGCGCTTACCGGCGCGAGCTGATCTCGGATCC 35519
QY 304 GGGTTCGCGC-----CGGCGCGAGGAGACGGGTGAGGAATCGGATCTGCGCGGCGG 357
Db 35520 CAGGTTCGCGGTTTGGCGGACCGCGGACCACTGGAACGCTTGGAGGAGACTGCGCGG 35579
QY 358 CT-----GCGTGAATGCGGAGCGGAGCGGATGTCAGGATGTCGCGCGGCTGTCGA 411
Db 35580 GCGTTCAGCGCGCGCTTCAACCGCGGAGGAGGACAAAGTCTGTGATGACTGGTCCG 35639
QY 412 GCGGAGTGGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 471
Db 35640 ACGGTGGCGGCGGAGTCTCGGTCAACGAGATCGCGGCGGCGGCGGCGGCGGCGG 35699
QY 472 TTCTGTAACCTGGATTCGATCTTCAATGACCGCGCGTGAACCTGCGGAAACCGGCTGCGGCG 531
Db 35700 TTCTGGAACCTGCGGCTTGAATCTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 35759
QY 532 GTGACCGGAGTCCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 591
Db 35760 GCGACCGGACTCTGATCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 35819
QY 592 ACCG 595

Db 35820 GCGC 35823

RESULT 10

US-08-804-198-1

Sequence 1, Application US/08804198
Patent No. 5945320

GENERAL INFORMATION:

APPLICANT: Burett, Stanley G.

APPLICANT: Kuhnstoss, Stuart A.

APPLICANT: Rao, Nagaraja R.

APPLICANT: Richardson, Mark A.

APPLICANT: Roelck, Paul R., Jr.

TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSER: PAUL R. CANTRELL 1138

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh

OPERATING SYSTEM: Macintosh 7.0

SOFTWARE: Microsoft Word 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,198

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CANTRELL, PAUL R.

REGISTRATION NUMBER: 36,470

REFERENCE/DOCKET NUMBER: P9113

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3885

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 44377 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 350..14002

FEATURE:

NAME/KEY: CDS

LOCATION: 14046..20036

FEATURE:

NAME/KEY: CDS

LOCATION: 20110..31284

FEATURE:

NAME/KEY: CDS

LOCATION: 31329..36071

FEATURE:

NAME/KEY: CDS

LOCATION: 36155..41830

US-08-804-198-1

Query Match 30.3%; Score 182.4; DB 2; Length 44377;

Best Local Similarity 60.1%; Pred. No. 8.3e-26;

Matches 363; Conservative 0; Mismatches 226; Indels 15; Gaps 3;

QY 7 TAGCGGCGGCGGCAACGCTTCTCGACGCGCTTCCGCGGCTGCGGCGCAAGTGGGCTG 66
DB 35220 TAGCGGCGGCGGCAACGCTTCTCGACGCGCTTCCGCGGCTGCGGCGCAAGTGGGCTG 35279
QY 67 CCGGTACCTTCATCGCTTGGGCTGCTGGGCGCGGCAAGACATGCGCCGGTACCGAGGCG 126
DB 35280 CCGGTACCTTCATCGCTTGGGCTGCTGGGCGCGGCAAGACATGCGCCGGTACCGAGGCG 35339

QY 127 GGC---GACTACCTTCGCGACGCCAGGCGCTGCGGCCCATGGACCCGAGCGGGCATGAG 183
DB 35340 GTTCACGGAATTTACACCCGAGCGCGGCTTCCGCCCATGGGCGCCGAGTCCGGCATGAG 35399
QY 184 GACTTCGCGACCACTTCGACCGCGGGAACCCGTGGGTGTGGTGGTGGACCTGGACCGG 243
DB 35400 GCACTGCAACCGCACTGAAACGAGGCGGACACCTGCGTCACTGCGCCGACATGACTGG 35459
QY 244 GAGCGGTTTGTGCAATCTTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303
DB 35460 GAACACTTGTGTCACCGGTTTCAACGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCG 35519
QY 304 GGGGTCGCGCGC-----CGGGCGGAGAGACCGGTCAAGATTCGATTCGCGCGCGCG 357
DB 35520 CAGTCCGCGCGGTGGGCGACGCCCGAACCCAGCTTGGAGCGCTTGGAGCGACTGCGCG 35579
QY 358 CT-----GGCGTCAATGCGGAGCGGCGGACGCTCAAGACATGTCGCGCGGTGTCGA 411
DB 35580 CGGTCGACGCGCGCGCTTCAACCGCGCGGAGCGGACCAAGTCTGTGTCGACTGTCGCG 35639
QY 412 GCCGAGTGGCAGCGGTGCTGGGCGCACGCGGACCGGATTCGAGCTGACGTCGCC 471
DB 35640 ACGGTGGCGCGGAGGTCTCTCGGTCAACGCGGATCGGCGGATTCGCGGACGACGAGG 35699
QY 472 TTCGATGACCTGGGATTCGACTTCATGACCGCGCGTGCAGCTTGGGAAACCGGCTCGG 531
DB 35700 TTCGGAACCTTCGACTTCGACTTCGCGCGGCGGTGCGGATGCGGCGCGCGCGCGG 35759
QY 532 GTGACCGGCGGCTCGGCTGGCGCACGACCATGCTTTCGACCAACCGGACAGTGGACCG 591
DB 35760 GCGACCGGACTGTAATGCGCGCGGACGCGTCACTTTCGACCAACCGGACAGTGGAC 35819
QY 592 ACCG 595
DB 35820 GCGC 35823

RESULT 11

US-08-804-227C-1

Sequence 1, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.

APPLICANT: Kuhnstoss, Stuart A.

APPLICANT: Roelck, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSER: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII (DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 43280 base pairs

```

1 TYPE: nucleic acid
2 STRANDNESS: single
3 TOPOLOGY: linear
4 MOLECULE TYPE: DNA (genom
5
6 FEATURE:
7 NAME/KEY: CDS
8 LOCATION: 816..14234
9
10 FEATURE:
11 NAME/KEY: CDS
12 LOCATION: 1451..19945
13
14 FEATURE:
15 NAME/KEY: CDS
16 LOCATION: 20010..31199
17
18 FEATURE:
19 NAME/KEY: CDS
20 LOCATION: 31232..35067
21
22 FEATURE:
23 NAME/KEY: CDS
24 LOCATION: 36249..41776
25
26 JS-08-804-227C-1

```

Query Match	29.7%	Score	172.6	DB 2	Length	43280
Best Local Similarity	57.3%	Pred.	No. 5.7e-24			
Matches 335, Conservative	0	Mismatches	244		Indels	6
					Gaps	1

QY	2	CTCTCTAAGCGCGGGGCAAGCGCTTCTCTGAGCGCCCTTCCGCGTGGTGGCGCACTG	61
Db	35127	CCGGCTTACGGGGGCGGCAAGCGCGCGCTGTGAGCGCCCTCGCGAGCGGCGCTTGGCGCG	35186
QY	62	GGCTGCGGATCACTCTCATGCGCTGAGGGATGTGAGGCGCGGCGAGAACATGACCGGTACCG	121
Db	35187	GGCTGCGCGGCACTCTCGATCGCTTGGGGGCTGTGAGGCGCGGAGGCAATGGGGAGGGGG	35246
QY	122	AGGGCGGAGCTACCTTGCGAGCGGAGGCTTGCGCGCATGGAACCGGACGCGGGGATCG	181
Db	35247	ACGGTGAAGAACTTCTTGAGCGGCGCGGAGCTTGCGGTGATGACCGCGGAGAGACGCGCTGG	35306
QY	182	AGGAGCTGCGGACCAACCTGTGAGCGCGGGGAGCCGTAAGGTGATCGGTGAGGACTTGAAAC	241
Db	35307	AAGCCCTGAGACCGGAGCCTCTGAGACGGGAGAGACACACCGTCTGTGAGGCGGATGTGACTT	35366
QY	242	GGAGAGCGATTGATGGAACCTGTTCAACCGCGCGCGCGCGCGGACCTCTTTCAGAGAACTCG	301
Db	35367	GGGAGCGGTTTGCCCGCGGCTTTCACCGCGTTTCCGAGCCAGATGAGCTGATCTCCCGCTGG	35426
QY	302	GTGGGGTCCGCGCGGGGGCGG-----AGAGAGCGGATCAGGAATGGATTTGCGCCCGGCG	355
Db	35427	TCTCGGACGCGCGGAGAGCGCGGGGGGAGAGAGCGCCCGGAGCGCAGCGCTGTTTGCACCGCG	35486
QY	356	GGCTGCGGCTCATGATCCGAGAGCGGCAAGTACAGAGCATGTGCGCCCGGCTGTGCGAGCGG	415
Db	35487	GGTTTGCGCGGCGCGCGCGGCACTGAGAGCGGCAAGAGATGTGCTTCCGCTGTGATCCGCGGCG	35546
QY	416	AGGTGGCAGCGGTGTGTGGGACACGGGCAACGCGGAGGTGATCGAGCGTGAAGTTCGCTTCC	475
Db	35547	ATGTGGCGCGCTGTACTCGGCCAACCCCGGGGAGACCGGAGACATTCGTCCTGAACCGTCTTCA	35606
QY	476	GTGACTTGAGATTGACTTCATGACCGCGGTGCACTTGCGGAAACCGGCTTCGCGGCGGTGA	535
Db	35607	AGGAGCTGGGGTTTCAGTTCCGCTTCACCGCGGTGAGCTGCGCGGCGGCTTGGCGCGGAGT	35666
QY	536	CGGGGGTCCGGGTGCGACAGACATGCTTTCGACACCCGAGCAG	580
Db	35667	GGGAGCGAAAGCTGCGCGCGAGCGCTGATCTTTCAGACATCCGACTCG	35711

```

? APPLICANT: Xue, Y.
? APPLICANT: Zhao, L.
? TITLE OF INVENTION: DNA encoding methymycin and pikromycin
? FILE REFERENCE: 600.438U1
? CURRENT APPLICATION NUMBER: US/09/105,537A
? CURRENT FILING DATE: 1998-06-26
? NUMBER OF SEQ ID NOS: 43
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 32
? LENGTH: 11220
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
US-09-105-537-32

```

Query Match	28.7%	Score 172.2;	DB 3;	Length 11220;
Best Local Similarity	57.3%	Pred. No. 6,8e-24;		
Matches 340; Conservative	0;	Mismatches 238;	Indels 15;	Gaps 1;

[illegible]

```

RESULT 12
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43

```

```

: SOFTWARE: FastSQ for Windows Version 3.0.
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-5

```

Query Match	28.7%	Score 172.2;	DB 3;	Length 36778;
Best Local Similarity	57.3%	Pred. No. 6.8e-24;		
Matches 340;	Conservative	0;	Mismatches 238;	Indels 15;
				Gaps 1;

QY	1	GGCCCTCAACGGGGGGGCAAGCCCTTCGACAGCCCTTCGCGCTGATGATGGGCGCACT	60
Db	19618	GGCCGTAACGGCCGGGTACGGGCTTCTCGAGCCCTCGCGGTGAGCACCGGGCGAC	19677
QY	61	GGGCTGCGGATCACTTCGATCGCTGGGGGTCTGTGGGCGGGCAGAACATGGCCGGTACC	120
Db	19678	GGCCCAACGGTGAACCTCGATGGCTGGAGCCCTGGGAGGGCAGCCGCTCACCAAGGT	19737
QY	121	GAGGGCGGGCACTACCTGGCGAGCCAGGGCTCGCGCCCATGGACCCGACGGGGGATC	180
Db	19738	GCGACCGGGGAGGGGCTGCGCCGCTTCGGGCTTGCGCCCTCGCCCGCCGACGGGGCTC	19797
QY	181	GAGGAGCTGGGAGCAACCCGTGAGAGCGCGGGGACCCGGGGTGTGGGTGTGACCTGGAC	240
Db	19798	ACCGCCTTGACACCGGGCTCGGCGACAGGGAGACCGCGGTCAAGATCCGACGTGCAC	19857
QY	241	CGGAGAGGATTGATGCAACTGTTCAACGCGCGCGCGCGCGCGCCCTCTTCGACGAACTC	300
Db	19858	TGGTCAGACTTTCCGCCCGGCTTCAACAAGGCCCGGCGACCTCTTCGCCGATCTG	19917
QY	301	GGTGGGGTCCGCGCCGGGGCCGAGGA-----GACGGGTGAGGAATTCGAT	345
Db	19918	CCCGAGGCGCGCGCGCGCTCGACGAGCAGACTCGACAGCGGCGCGGACGACACGTC	19977
QY	346	CTGCGCCGGGGGCGCTGCGATCGATGCGGAGCGCGGAACCTCAAGAGATGTCCCGGCTG	405
Db	19978	CTGAGCCGCGAGCTCGGTGCGCTCAACCGGGCGGAACAGACGGCGGTATGACGAGTTG	20037
QY	406	GTCGAGCCGAGGTGGCAGCGGTGCTGGGCGACAGGACCGCACGATGATCGAGGATGAC	465
Db	20038	GTCGCGGAGGACCTCGCGGTGGTCTCAACCAACCTCTCCCGGAGGCCGTCAACAGGGG	20097
QY	466	GTGCGCTTCCGTGACTCGGAGTTGCACTCTCAATGACCGCGGTGACTTGGGAACTGGCTC	525
Db	20098	CGGCGCTTCCGTGACTCGGATTCGACTCGTGAACGGCGGTGAGGCTTCGCAACCGCTC	20157
QY	526	GCGGCGGTGACCGGGGGTCGGGGTGGCGACAGCAACATGTCTTTCGACCAACCCGAC	578
Db	20158	AAAGACCCACCGGCGCTGCGCCCTCCGGGCACTGTGCTTTCGACTCAACCGAC	20210

US-09-320-878-19
Sequence 19, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320, 878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119, 139

```

? EARLIER FILING DATE: 1999-02-08
? EARLIER APPLICATION NUMBER: 60/100,880
? EARLIER FILING DATE: 1998-09-22
? EARLIER APPLICATION NUMBER: 60/087,080
? EARLIER FILING DATE: 1998-05-28
? NUMBER OF SEQ ID NOS: 34
? SOFTWARE: Patentn Ver. 2.0
? SEQ ID NO: 19
? LENGTH: 38506
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

Query Match	28.7%	Score 172.2;	DB 3;	Length 38506;
Best Local Similarity	57.3%;	Pred. NO. 6.8e-24;		
Matches 340;	Conservative 0;	Mismatches 238;	Indels 15;	Gaps 1;

OY 1 GCGCTTCAACGCGGGGCAACGCGCTTCTCGAAGCCTTGAGCCGTGATGATGGGCGCACT 60
 Db 17760 GGGCGCTAACCCGCGCGGTACGCGCTTCTTGAGCGCTTGGCGGTGAGCACGGGCGGAC 17819
 OY 61 GGGCTTCGCGGTCACTTGATGCGCTTGAGGCTTGTGGGCGGGCGAGAACATGCGCGGTACC 120
 Db 17820 GGGCCCAACCGGTGACTCTGGGTGGCTTGAGAGCCCTTGGAGAGGAGAGCGCGCTCACACAGGCT 17879
 OY 121 GAGGGCGGGCGACTACTGTGCGGAGACCGAGGCGCTGCGCGCATGAGCCCGAGCGGGGATC 180
 Db 17880 GCGACCGGGGAGGCGGTGCGCGCGCTCTGGCGCTGCGCGCCCTTGGCGCCCGCGACGGGCGTCT 17939
 OY 181 GAGGAGGTGCGGACCAACCTCGAGCGGCGGGGAGCCGTGGGTGTCCGTGTGTGACCTTGAC 240
 Db 17940 ACCGCGCTTGACACCGCGCTCGGCGACCGCGAGACACCGCGGTACATGCGCTCACTCGAC 17999
 OY 241 CGGAGGCGGTTTCGTGCAGCTGTTCACCGCGCGGCGCGCGCGCGCTTCTTGACGAACTTC 300
 Db 18000 TGGTTCAGACTTTCGCGCGCGCGCTTCAACAGCGCGCGCGGACCTCTCTCGCGCATCTG 18059
 OY 301 GGTGGGGTCCGCGCGCGGGGCGAGGA-----GACCGGTCAAGAAATCGAT 345
 Db 18060 CCGAGGCGCGCGCGCGCGCTCGACGAGCAGTGCAGCAGCGCGCGCGACGACACCGTC 18119
 OY 346 CTGCGCCCGGCGGTGCGGTGCATATGCGCGGAGCGGAACTGACAGACATGTGCGCGCGCTG 405
 Db 18120 CTAGCGCGGAGGTCTGCTGTCGCTCAACCGCGCGCGGAAACAGACCGCGCTGATGAGGANGTGG 18179
 OY 406 GTCCGAGCCGAGGTGTCAGCGGTGCTGGGCGACAGGACGCGCGAGCTGATCGAGGTGAC 465
 Db 18180 GTCCGAGAGCACTTGCGCCTGTGTCTCAACCAACCCCTCCCCGAGGCGGTGCACAGGG 18239
 OY 466 GTGCGCTTCCGTGACTCTGGGATTTGACCTTCATGACCGCGGTGACTGCGAGACCGGCTC 525
 Db 18240 CGGCGCTTCCGTGACTCTCGGATTTGCACTGCTGAGCGGCGGTGACTCGCAACCGGCTC 18299
 OY 526 GCGGCGGTGACCGGGGTCGCGGGTGGCGACAGACATGTCTTTGACACCCGAC 578
 Db 18300 AAGAGCGCAACCGGCTGCGCGCTTCCCGGCACTGTGTCTTGCATTACCGAC 18352

```

RESULT 15
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polypeptide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 2310.57 Seconds
(without alignments)
12149.485 Million cell updates/sec

Title:	US-10-611-442-2_COPY 1_600
Perfect score:	600
Sequence:	1 ccgcgcctgcgcgagcgtac.....gtcccccgcgcgctgggtg 600

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database :

```

1:  gb_esc1: *
2:  gb_esc2: *
3:  gb_esc3: *
4:  gb_hlc: *
5:  gb_esc4: *
6:  gb_esc5: *
7:  gb_esc6: *
8:  gb_esc7: *
9:  gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.4	13.6	925	10	CNS0091P	AL053013 Drosophi
2	80.2	13.4	1438	9	B2557931	B2557931 pasci-60
3	79.2	13.2	1143	5	BU502934	BU502934 AGENCOURT
4	75.6	12.6	888	10	AG030591	AG030591 Pan trog
5	75.2	12.5	925	10	CNS0091P	AL053013 Drosophi
6	74.8	12.5	935	10	CNS0065XK	AL066051 Drosophi
7	73	12.2	1169	5	BQ922913	BQ922913 AGENCOURT
8	72.8	12.1	852	10	AG060114	AG060114 Pan trog
9	72.8	12.1	985	6	CA981964	CA981964 AGENCOURT
10	72.6	12.1	1003	10	AG159111	AG159111 Pan trog
11	72.2	12.0	1469	10	AG365356	AG365356 Mus musc
12	71.8	12.0	1152	10	AG076818	AG076818 Pan trog
13	71.6	11.9	996	10	AG072414	AG072414 Pan trog
14	71.6	11.9	1101	10	AG039543	AG039543 Pan trog
15	71.6	11.9	1440	10	AG332984	AG332984 Mus musc
16	70.8	11.8	935	10	CNS0065XK	AL066051 Drosophi
17	70.4	11.7	492	7	C0851644	C0851644 IM SH5_00
18	70.4	11.7	519	7	C0825295	C0825295 IM SH5_00
19	70.4	11.7	536	7	C0851640	C0851640 IM SH5_00
20	70.4	11.7	581	7	C0834687	C0834687 IM SH5_00
21	70.4	11.7	532	7	C0851643	C0851643 IM SH5_00
22	70.4	11.7	597	7	C0825301	C0825301 IM SH5_00

23	70.4	11.7	609	7	CO825285	CO825285 LM GB5_00
C 24	70.4	11.7	1122	3	BM914385	BM914385 AGENCOURT
C 25	70.4	11.7	1569	10	AG341503	AG341503 Mus muscu
C 26	70	11.7	876	8	DN477453	DN477453 Alt1033X
C 27	69.8	11.6	743	10	CM951338	CM951338 TC836_4
C 28	69.8	11.6	937	10	AG080291	AG080291 Pan trogl
C 29	69.8	11.6	1134	10	AG073255	AG073255 Pan trogl
C 30	69.8	11.6	1552	10	AG430101	AG430101 Mus muscu
C 31	69.6	11.6	1635	10	CL9777013	CL9777013 OrlFC043
C 32	69.4	11.6	2400	10	CL984281	CL984281 OrlFSC047
C 33	69.2	11.5	1134	3	BM915656	BM915656 AGENCOURT
C 34	68.8	11.5	564	7	CO856522	CO856522 LM S15_00
C 35	68.8	11.5	584	7	CO851638	CO851638 LM_S15_00
C 36	68.6	11.5	1856	10	AG072699	AG072699 Pan trogl
C 37	68.6	11.4	932	10	CNS007220	AL067642 Drosophi1
C 38	68.4	11.4	1267	10	CL476590	CL476590 SAIL_258
C 39	68.4	11.4	2243	10	AG381986	AG381986 Mus muscu
C 40	68.2	11.4	952	5	B0879171	B0879171 AGENCOURT
C 41	68.2	11.4	1713	10	AG435689	AG435689 Mus muscu
C 42	68.2	11.4	1947	10	CL970284	CL970284 OrlFC041
C 43	68	11.3	962	5	BUS40148	BUS40148 AGENCOURT
C 44	68	11.3	1273	3	BM562099	BM562099 AGENCOURT
C 45	67.8	11.3	640	11	LB4F046709	BX543344 Telistmani

ALIGNMENTS

RESULT 1	CNS00091P	925 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	CNS00091P				
DEFINITION	Drosophila melanogaster genome survey sequence T8T3 end of BAC # BACR19D16 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL053013				
VERSION	AL053013.1	GI:4934461			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 925)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FR:NCRC (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuoto Osegaawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and B8T libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
SOURCE	1..925				
	/organism="Drosophila melanogaster"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:7227"				
	/clone="BACR19D16"				
	/clone_11b="RPc1-98"				
	/note="end : T8T3"				
ORIGIN					

	Query Match	13.6% Score 81.4 DB 10; Length 925;
	Best Local Similarity	13.2%; Pred. No.2.2e-06;
	Matches	52; Conservative 201; Mismatches 140; Indels 0; Gaps 0;
OY	145 CCGGACAGATGAGCTTACCAGCTGCCTTGACACCCCGTAGACTCGAGCGATCTCC	204
Db	533 SSMCKKCGSTTBGSITTTTSSGSGYGKCCSBSGBSCSCBSCSCCBBCCC	592
OY	205 TGCGCCGACGGGTCTGTGTGTGAACGGAGCGGCACTCCCCCGCTCGAACGACGTG	264
Db	593 CCSCSBTCSSSSBSSSKCSSTSBCSCCSSKSVCGRSCSSSSSCSSSSSTSSSTSS	652
OY	265 TCCCGACGCGCTTGAAACAGCGCGGCGGACCGCTGCTTTGTGACCGCGAGTCGCG	324
Db	653 TSXSSBSGSSSSSSSYTTTSKTASBGSGMSAAGGSGSSTSTSSSSSSSTSSSVS	712
OY	325 CCCGATTCGCGCGCACCTGCACGCCCGTGCAGCGCACCGCCCTGTTCACGTGCTCTC	384
Db	713 SGSRKSTBSBSBSBSBSGSSSSSTSSBSLCTSTSSSSSYBSSTCS.CTCCCSYSBS	772
OY	385 TGCTTCGCGCTGACCGAGCGCGATGTGTGCAGACACCCGAGCTGACACCCCTCGGTGG	444
Db	773 TSSSSSTBSWGSTSGSSSSSVSTSSSDSTTCCSCCTCMCTCTCTBMCTTSSGSS	832
OY	445 TCACAGCGCTTCGCGCGCACCGGATTCAGCTGCCCTTGTGCTGTGTGACAGGACGCG	504
Db	833 SSGKGGVTKCGCGCGCGSSSTMGBGTSAACSSSSSCSSSVSSBSKSSABSSSVSSG	892
OY	505 CCGCCGTGACCGCTCGGAGACGACGTTCATCCGG	537
Db	893 SSGVSNSSSASKSSSGSVSGSGSGGSGSVG	925

RESULT 2			
LOCUS	BZ557931	1438 bp	DNA
DEFINITION	pacsl-60_6778.x25p11 pacsl-60 <i>Pseudomonas aeruginosa</i> genomic clone		
ACCESSION	pacsl-60_6778, genomic survey sequence.		
VERSION	BZ557931		
KEYWORDS	BZ557931.1	GI:27171396	
SOURCE	GSS.		
ORGANISM	<i>Pseudomonas aeruginosa</i>		
REFERENCE	<i>Pseudomonas aeruginosa</i>		
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
TITLE	1 (bases 1 to 1438)		
JOURNAL	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haetings,M., Burns,J.L., Kaul,R. and Olsen,M.V.		
COMMENT	Whole-Genome-Sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library J. Bacteriol. (2002) In press		
	Contact: Chris K. Raymond		
	Genome Center		
	University of Washington		
	Box 352145, Seattle, WA 98105-2145, USA		
	Tel: 2062216954		
	Fax: 2066857244		
	Email: craymond@u.washington.edu		
	Class: shotgun.		
FEATURES	Location/Qualifiers		
source	1..1438		
	/organism="Pseudomonas aeruginosa"		
	/mol_type="genomic DNA"		
	/strain="1-60"		
	/db_xref="taxon:287"		
	/clone_lib="pacsl-60_6778"		
	/clone_lib="pacsl-60"		
	/note="Clinical isolate 1-60 Whole genomic shotgun library."		
ORIGIN			
Query Match	13.4%	Score 80.2;	DB 9;
Best Local Similarity	47.8%;	Pred.No. 3.7e-06;	Length 1438;

[illegible]

```

RESULT 3
BUS02934
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BUS02934 1143 bp mRNA linear EST 12-SEP-2002
AGNCOURT 8929661 NIH_MGC_94 Mus musculus CDNA clone IMAGE:6489733
5', mRNA sequence.
BUS02934 BUS02934.1 GI:22808397
EST.
Mus musculus (house mouse)
Mus musculus
Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
1 (bases 1 to 1143)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: gcapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLNL at:
http://image.lnl.gov

```


Df | ||
DB 858 CANNCCCA 865

RESULT 5
CONS0091P/c

LOCUS CENS0091P 925 bp DNA linear GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC # BACh1916 of RPCT-98 library from Drosophila melanogaster ('fruit fly'), genomic survey sequence.

ACCESSION AL053013

VERSION AT053013.1 GI:4934461

KEYWORDS GSS.

SOURCE Drosophila melanogaster ('fruit fly')

TAXONOMY Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Epitheroidea, Drosophilidae, Drosophila.
1 (bases 1 to 925)

REFERENCE Genoscope.

AUTHORS Direct Submission

TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefigenoscope.cns.fr

JOURNAL Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with The Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mammoset in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw ap, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the DACPC Resource Center can be found at http://dacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source location/Qualifiers

1..925 /organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACH1916"
/clone_1lb="RPCT-98"
/note="end : TERT"

ORIGIN

Query Match 12.5%; Score 75.2; DB 10; Length 925;

Best Local Similarity 13.3%; Pred. No. 3.8e+05;

Matches 52; Conservative 194; Mismatches 144; Indels 0; Gaps 0,

QY	CGGATGAGAGTCATCCCTCTGTGGCTGTGTGACCAAGGAGACGGCCGCGCCGTGACGTCGAGGA	523
Db	621 GGAGGVSASSGAGSSSVSSSGAGSSGGGGGAGVAGAGSSSSGSSGSSGAGSVCCSGCSGCGM	562
QY	524 CGACGTGCATCCGGCCGACGACCATGTGTCG	553
Db	561 CCGCCSSAAAAAASCTVAASCGMCGKSG	532
RESULT 6		
CNS006XK/c		
LOCUS		
DEFINITION		
Drosophila melanogaster genome survey sequence T7 end of BAC #		
BACRI4N09 of RPCI-98 library from Drosophila melanogaster (fruit		
fly), genomic survey sequence.		
AL066051		
AL066051.1 GI:4945019		
GSS.		
Drosophila melanogaster (fruit fly)		
Drosophila melanogaster		
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
Ephydroidea; Drosophilidae; Drosophila.		
1 (bases 1 to 935)		
Genoscope.		
Direct Submission		
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
- Web : www.genoscope.cns.fr)		
Determination of this BAC-end sequence was carried out as part of a		
collaboration with the Berkeley Drosophila Genome Project (BDGP).		
The BDGP is constructing a physical map of the Drosophila		
melanogaster genome using these BACs. For further information		
please see http://www.fruitfly.org The BDGP Drosophila		
melanogaster BAC library was prepared by Kazutoyo Osoegawa and		
Aaron Mammoser at the Roswell Park Cancer Institute in Buffalo,		
NY. The library is named RPCI-98 and was constructed by partial		
EcoRI digestion of Drosophila DNA provided by the BDGP from the		
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's		
P1 and EST libraries. A more detailed description of the library		
and how to order individual BAC clones, the entire library, or		
filters for hybridization from the BACPAC Resource Center can be		
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .		
Location/Qualifiers		
1..935		
/organism="Drosophila melanogaster"		
/mol_type="genomic DNA"		
/db_xref="taxon:7227"		
/clone="BACRI4N09"		
/clone_11b="RPCI-98"		
/note="end : T7"		
FEATURES		
source		
ORIGIN		
Query Match 12.5%; Score 74.8; DB 10; Length 935;		
Best Local Similarity 30.2%; Pred. No. 4.6e-05;		
Matches 125; Conservative 114; Mismatches 172; Indels 3; Gaps 1		
QY	112 GCGTCCCGCTCCCTCTGGGCGGGTCCCGACACCGGAGACGAGTGGCGTTACAGCTCG	171
Db	906 GSCCCSSGSCCGSCGCGCCSSGSSSSCCSSSCGSCCGCCSGSSSCSGCGGCGSGS	847
QY	172 CTTGACACCCCGTGCACCTCGGCGGTCCTCTCTGAGCGAGACGAGTCTCTGTGTGACCG	231
Db	846 SSGCGCCSSGSCCGCCSSCGSSCGSCSSCTCGGCGCCSCGCGCCCGGCGCGSGCG	787
QY	232 GAGCGGAGTACCCCGGCGCTTGACGAGACGTGTCTCGGACGCGCTTGAAACAGCGCGG	291
Db	786 SSGCGGCGGCGCCSSCGCGSKCGCGCGSGGSGSGSSGSGSGSGSGSGSGSGSGSGS	727
QY	232 GACCGTGTGTGTGTGACACCGGACGAGTGGCGCCCGGATCGGCGCCGACATCGACGCG	351
Db	726 CGCGSSGCGCGCGCG---CGGCGCGCGCGSGCGSGCGCGCGCGCGCGCGCGCGCG	670

OY	TCGAGGGACCGCCCTGTGCATGTAGTCTCTCTCTCGCGCTGCCAGAGGGCGGTCTG	411
Dd	CCCCSSSSCASSCSLSSSSSCSSSCSSSgcccgcgccscgcgcgcscccgccscccccg	610
OY	TGCAGCACCCCAAGCTTGACACCTTCGCGTTGTGTGCAGGCGCTCGGCGCAGCGGGATGC	471
Dd	CAGCSCAAGCSCSGMAGVAGSGASRVSLSGSSSSGLSASGCGCGCCAGCGRASGRKMSAGS	550
OY	ACGTCCCCCTGTGTGCTGTGAACAAGAGACGCCGCGCGTGACCGTTCGGAGACG	525
Dd	SGRCGGAAGSGSGGSSSKRKAGAAAGCAGCSATYKSGSCCGCBKKHGAACAS	496

RESULT 7	
B0922913	
LOCUS	B0922913
DEFINITION	1169 bp mRNA linear EST 20-AUG-2002
ACCESSION	AGNCOURT 8922123 NCI CGAP Co24 Mus musculus cDNA clone IMAGE:6396111 5' , mRNA sequence.
	D000000000

ACCESSION	BQ922913
VERSION	BQ922913.1
KEYWORDS	GI:22337944
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
	Mus musculus

REFERENCE	1 (bases 1 to 1169)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Tissue Procurement: The Cepco Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LLAM1381 row: n column: 16
 high quality sequence stop: 147.

FEATURES	Location/Qualifiers
source	1. .1169

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6396111"
/lab_host="DH10B (T1 phage-resistant)"
/clone_id="NC1_GCAP_Coz4"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unirectionally. Primer: Oligo dT Average insert size 1.6 kb. A Constructed by Life Technologies. Note: this is a NC1_GCAP library."

```

ORIGIN

Query Match	12.2%	Score 73;	DB 5;	Length 1169;
Best Local Similarity	46.7%	Pred. No. 0.00011;		
Matches 279; Conservative	0;	Mismatches 313;	Indels 6;	Gaps 2

[illegible]

QY 182 GGTGACCTCGGGAGCGTCTCTCTTGGCGGAGAGGGGCTGTGGTGGTATACCGAGACGAGT 24.1
 Db 382 GGGCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 44.1
 QY 242 ACCCGCGGCGCTGAGCGAGCGTGGTTCGCGAGCGGCTGTGAAAGCGCGGGCGGACCGTCT 30.1
 Db 442 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 50.1
 QY 302 GTTGTGC--ACGCGCGAGTCCGCGCGCGCGCGGATCCGCGCGCGCACTGACCGCGGTGACGCG 35.5
 Db 502 GAGCGGCGGCGG 56.1
 QY 360 ACCGCGCGTGCACATGGTGTCTCTGTGTCGCGCGCTCGCGAGAGGAGGAGTGTGTGACGAC 41.9
 Db 562 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 62.1
 QY 420 CCCAGCGCTGAGCAACCTTCGCGTGTGGTTCAGAGCGCTTCGCGCGAGCGCGGAAATCGACGTCCCG 47.9
 Db 622 GGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 67.7
 QY 480 CTGTGGCTGTGTGATCCAGAGAGCGCGCGCGCGCGCTGTGACCGGTGAGAGAGAGCTCGATCCGCGCC 53.9
 Db 678 GCCGAGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 73.7
 QY 540 CAGGCGCATGTGTGCGGTGGGCTCGCGCGGGGAGTGGTGGGCGGTGAGATCCCGCGCGCGGTGTCG 59.7
 Db 738 CGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 79.5

RESULT 8	AG060114/c	852 bp	DNA	linear	GSF 03-NOV-2001
LOCUS	AG060114				
DEFINITION	Pan troglodytes DNA, clone: FIB-047705.F, genomic survey sequence.				
ACCESSION	AG060114				
VERSION	AG060114.1	GI:1661344			
KEYWORDS	GSF.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes				

REFERENCE	1
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	BAC end sequences of Library PTB
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 852)
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submissions
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of

COMMENT
Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

FEATURES

SOURCE

```

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-047105.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

```

ORIGIN

Query Match 12.1%; Score 72.8; DB 10; Length 852;
Best Local Similarity 46.0%; Pred. No. 0.00012;
Matches 267; Conservative 0; Mismatches 308; Indels 6; Gaps 1;

QY 1 CCGGCTCGCCGAGCGTACACCCGGGGGTGAGGTGACCTGCGGTACCGCAGTGGGTG 60
DB 831 CCGGCTCGGCG 772
QY 61 AGGAGCGCGCGGTGACCTGCGGTGACCGGTTCACACGACAGACTTCTGCTCCCG 120
DB 771 CG 718
QY 121 TCCGCTGAGCGCGGTGCG 180
DB 717 GCG 658
QY 181 CCGTGCACCTGCGCGCGGTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 657 GCG 598
QY 241 TACCGCGCGCGGTGACCGGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 597 CCGGCG 538
QY 301 TGTGTGACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 537 GCG 478
QY 361 CCGCGCTGTGACCTGTGTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 477 CCG 418
QY 421 CCAAGCTGGAACACCTGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 417 CCG 358
QY 481 TGTGACTGTGACCG 540
DB 357 CCG 298
QY 541 AGGCGATGTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
DB 297 CCG 257

RESULT 9
CA981964 985 bp mRNA linear EST 27-FEB-2003
DEFINITION AGNCOURT 11280952 Wellcome CRC PSK egg Xenopus laevis cDNA clone
IMAGE:6870527 5', mRNA sequence.

ACCESSION CA981964
VERSION CA981964.1 GI:27514618
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM

REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@bcr-remail.nih.gov
Tissue Procurement: N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
Gordon (Wellcome/CRC Institute)
cDNA Library Preparation: N. Garrett, P. Lemaire, A.M. Zorn, and
J.B. Gordon (Wellcome/CRC Institute)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: XGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.jnl.gov
Plate: LLAM4501 row: a column: 22
High quality sequence stop: 241.
Location/Qualifiers

FEATURES

source

1. 985
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6870527"
/issue_type="egg"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC PSK egg"
/notes="Vector: Bluescript SK-; Site 1: NotI; Site 2:
EcoRI; cDNAs were oligo-dT primed and directionally
cloned. Library was constructed by N. Garrett, P. Lemaire,
A.M. Zorn, and J.B. Gordon (Wellcome/CRC Institute)."
Note: This is a Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 12.1%; Score 72.8; DB 6; Length 985;
Best Local Similarity 46.9%; Pred. No. 0.00012;
Matches 250; Conservative 0; Mismatches 282; Indels 1; Gaps 1;

QY 35 GGTGACTGCGGTACCGCAGTGGGTGAGAGACCGCCGCGTGCACCTGCGGTTCACCGTT 94
DB 343 GCG 402
QY 95 CCAAGCAGACAACTTCTGCTCCCGGTCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 154
DB 403 GCG 462
QY 155 GTGCGCTTACGAGCTGCG 214
DB 463 GCG 522
QY 215 GGTCTGTGTGTGACCGCGAGCGGAGTACCCCGCGCTGACCGACGCTGTGCGCGACG 274
DB 523 GCG 582
QY 275 CTTGGAACAGCG 334
DB 583 CCG 642
QY 335 GCGCGCACTGACGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
DB 643 GCG 701
QY 395 GCGCGAGCGCGGTGCTGTGACGACCCGCGCTGAGACACCTGCTGTGCGCGCT 454
DB 702 CCG 761
QY 455 CCGCGCGCGCGGTGACGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 514
DB 762 CCG 821
QY 515 GGTGAGAGACGCTGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
DB 822 CCG 874

RESULT 10

AG159111

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AG159111 1003 bp DNA linear GSS 09-JAN-2002
Pan troglodytes DNA, clone: RP43-024116.T7, genomic survey
sequence.
AG159111
AG159111.1 GI:16688789
GSS.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]

RESULT	12
AG076818	
LOCUS	AG076818 1152 bp DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.
ACCESSION	AG076818
VERSION	AG076818.1 GI:16628620
KEYWORDS	GSS.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	Fujiyama, A., Hattori, M., Toyoda, A.,	Taylor, T. D., Yada, T.,		1-7-22	Suehiro-chou, Tsurumi-ku,	Yokohama, Kanagawa	230-0045, Japan
	Totoki, Y., Watanabe, H. and Sakaki, Y.						
	BAC end sequences of library PTB						
	Unpublished						
	2 (bases 1 to 1152)						
	Fujiyama, A., Hattori, M., Toyoda, A.,	Taylor, T. D., Yada, T.,					
	Totoki, Y., Watanabe, H. and Sakaki, Y.						
	Direct Submision						
	Submitted (02-AUG-2001)	Asao Fujiyama, The Institute of Physical					
	and Chemical Research (RIKEN)	Genomic Sciences Center (GSC) ;					
	1-7-22 Suehiro-chou, Tsurumi-ku,	Yokohama, Kanagawa	230-0045, Japan				
	(E-mail:chiumpes@sc.riken.go.jp,	URL:http://hgc.gsc.riken.go.jp/,					
	Te1:81-45-503-9111, Fax:81-45-503-9170)						
	Clones are derived from the chimpanzee BAC library PTB this BAC end						
	was generated during the Red process and may have higher chance of						
	clone tracking errors.						
	PRIMERS						

FEATURES	location/Qualifiers
source	1..1152
	/organism="Pan troglodytes"
	/mol_type="genomic DNA"
	/db_xref="taxon:9598"
	/clone="PTB-071C05.R"
	/sex="male"
	/cell_type="lymphoblast"
	/clone_lib="PTB Chimpanzee Male BAC library"
ORIGIN	

Query Match	12.0%	Score 71.8	DB 10	Length 1152
Best Local Similarity	48.4%	Pred. No. 0.00018		
Matches 237	Conservative 0	Mismatches 250	Indels 3	Gaps 2
QY	113	GCTCCCGATCCCTCCGAGCCGAGTCCCGACACCCGAGCAACGAGTGCCTTACCACTCCG	172	
Db	630	GGCGAGGAGGCGACMGCCGCCCCCGGCGCCCCCGGCGCCCCCGANNNGGCGAGCGCGCGCG	689	
QY	173	CTGGCACCCCGTGCACCTCCGAGCGAGTCTCTCCCTTGACCGGACCGGTCTTGATGAAACCG	232	
Db	690	CGGCGCGGCGCNCCCCCGGCGCCCCCGCCGCGCCCCCGCGCGCGCGCGCGCGCGCGCG	749	
QY	233	AGGCGGAGTACCCCGGCGCTGAGACGAGCGTGTCCGCGACGCGCTTGAAACAGCGCGGAGC	292	
Db	750	CGGCGACMGCGGCG	809	
QY	293	GACCGTGTGTGTGTGACACCGCGGAGTGCAGCGCCCGGAGTCCGCGCGCGCACTGCAGCGCGT	352	
Db	810	GCCTCGCGCGAGGCGGCG	869	
QY	353	CGACGCGACCGACCTCTGTCCACTGTGTCTCTGTGTGCGCTCGCCGAGGCGGATGTGT	412	
Db	870	CG	928	
QY	413	CGACGACCCCGACGCTGAGCACCCCTCGAGTGTGATCAGAGCGGTCCGAGCGGAGATCGA	472	
Db	929	GGCG	988	
QY	473	CGTCCCGCTGTGTGTGTGACCAAGAGACGCGCGCGCGCGTACCGTTCGAGACGACGTTCGA	532	
Db	989	CGGCG	1048	
QY	533	TCCGCGCCGAGCGCATGTGTCGTGGCGC--TGGCGCGGAGTGTGGGCGTGTGAATCCCCCGC	590	
Db	1049	GGCG	1108	
QY	591	CGGTGGGAGTG 600		
Db	1109	CGCGCGCGCGCG 1118		

RESULT 13
AG072414 LOCUS 996 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-063011.R, genomic survey sequence.
ACCESSION AG072414
VERSION AG072414.1 GI:16624216
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 996)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suichi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbeegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/)
Tel.:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rnd process and may have higher chance of
clone tracking errors.
PRIMERS

RESULT 15
AG332984/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
ORIGIN
Query Match 11.9%; Score 71.6; DB 10; Length 1440;
Best Local Similarity 47.0%; Pred. No. 0.0002;
Matches 254; Conservative 0; Mismatches 285; Indels 2; Gaps 1;
QY 1 CCGCGCTCGCGAGCGTACACCCGCGGCGTGGAGTCACTGCGGTACCGGAGTGGTG 60
DB 1062 CCCCCCTCG 1003
QY 61 AGGAGCGCCCGGTGCACTGCGCGGTCTACCGGTTCAACGACAGACTTCTGGCTCCGG 120
DB 1002 CGG--CGGCG 945
QY 121 TCCCGCTGGGCGCGGTCCCGGACACCGGCGAGAGTGGGTTACAGCTCGCTGGACAC 180
DB 944 CCG 885
QY 181 CCGTGAACCTCGGCGGTCTCTCGCGCGGACGCGGTCTGTGTGTGACCGGAGCGGCG 240

DB 884 CCG 825
QY 241 TACCCCGCGCTGAGACGAGCTGATCGGACGAGCGCTGAAACAGCGCGGCGGACCGCTGC 300
DB 824 GCG 765
QY 301 TGTGTGACACCGGCGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 764 CCG 705
QY 361 CCG 420
DB 704 CCG 645
QY 421 CCAAGCTGACACCGCTGCGGTGATCGAGCGCGCTCGGCGAGCGCGGATGACGTCGCC 480
DB 644 CCG 585
QY 481 TGTGTGATGACCGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
DB 584 CCG 525
QY 541 A 541
DB 524 A 524

Search completed: April 7, 2006, 14:04:00
Job time : 2314.57 secs

[illegible]

CC under conditions in which the cell expresses a product of a gene encoded
CC by the nucleic acid under conditions in which the unmodified polyketide
CC is present, and producing under conditions in which the cell produces
CC megalomycin, and can attach megalomycin to a polyketide, where the cell, it
CC its naturally occurring non-recombinant state cannot produce megalomycin.
CC The present sequence contains downstream megalomycin modification enzyme
CC genes.

SO Sequence 17596 BP; 2325 A; 5966 C; 6604 G; 2701 T; 0 U; 0 Other;

Query Match 100.0%; Score 601; DB 12; Length 17596;

Best Local Similarity 100.0%; Pred. No. 1,8e-91;

Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GCCCTCAACGCGGCGGGAAGCCTTCTCGACGCTTCCGCGCGTGGTGGCGGAGT 60
    |||
DB 1200 GCTCTCAACGCGGCGGGAAGCCTTCTCGACGCTTCCGCGCGTGGTGGCGGAGT 1259
QY 61 GGGCTGCGCGTCACTCGATCGCTGGGCTCTGTGGGCGGCGAGAACATGGCCGTACC 120
    |||
DB 1260 GGGCTGCGCGTCACTCGATCGCTGGGCTCTGTGGGCGGCGAGAACATGGCCGTACC 1319
QY 121 GAGGCGCGGCGACTATCTGCGAGCGGCGCTTGGCGCCATGAAACCGCGAGCGGCGATC 180
    |||
DB 1320 GAGGCGCGGCGACTATCTGCGAGCGGCGCTTGGCGCCATGAAACCGCGAGCGGCGATC 1379
QY 181 GAGGAGCTGCGGAGACCACTCGAGCGGCGGAGCCGTTGGGCTGTGGTGGTGGACCTTGAC 240
    |||
DB 1380 GAGGAGCTGCGGAGACCACTCGAGCGGCGGAGCCGTTGGGCTGTGGTGGTGGACCTTGAC 1439
QY 241 CGGAGCGGCTTCTGTGAACTGTTCAACGCGCGCGCGCGCGCGCCCTTCTGAGGAACTC 300
    |||
DB 1440 CGGAGCGGCTTCTGTGAACTGTTCAACGCGCGCGCGCGCGCGCCCTTCTGAGGAACTC 1499
QY 301 GGTGGGCTCGCGCGCGCGCGCGGAGGAGACCGGCTGAGAACTGCTGCGCGCGGCTG 360
    |||
DB 1500 GGTGGGCTCGCGCGCGCGCGCGGAGGAGACCGGCTGAGAACTGCTGCGCGCGGCTG 1559
QY 361 GCGTCTGATGCGCGGAGCGGAGCGTCAACGAGCATGTGCGCGCGCTGCTCGAGCGGAGT 420
    |||
DB 1560 GCGTCTGATGCGCGGAGCGGAGCGTCAACGAGCATGTGCGCGCGCTGCTCGAGCGGAGT 1619
QY 421 GCAAGCGGCTGCGGCGCGGAGCGGCGGAGCGGCTGATGAGCGTGTGCGCTTCCGTGAC 480
    |||
DB 1620 GCAAGCGGCTGCGGCGCGGAGCGGCGGAGCGGCTGATGAGCGTGTGCGCTTCCGTGAC 1679
QY 481 CTGGGATTCGACTCGATGACCGCGCTGCACTTGGCGGAAACCGGCTGCGCGCGTGAACCGG 540
    |||
DB 1680 CTGGGATTCGACTCGATGACCGCGCTGCACTTGGCGGAAACCGGCTGCGCGCGTGAACCGG 1739
QY 541 GTCCGGGTGGCGACGACCATGTCTTTCGACCAACCGGAGAGTGGACCGGCTCAACCGCGGAC 600
    |||
DB 1740 GTCCGGGTGGCGACGACCATGTCTTTCGACCAACCGGAGAGTGGACCGGCTCAACCGCGGAC 1799
QY 601 T 601
DB 1800 T 1800
```

RESULT 2

AAAF30757

ID AAF30757 standard; DNA; 47981 BP.

AC AAF30757;

DT 21-JUN-2001 (first entry)

DE Micromonospora megalomicea megalomiclin biosynthetic gene cluster.

XX Megalomiclin; meg gene; polyketide synthase; antibiotic; motilide;

KW antiparasitic; ds.

XX Micromonospora megalomicea subsp. nigra.

```
XX Key Location/Qualifiers
FT CDS complement(1..144)
FT /*tag= a
FT /gene= "megBVI(megT)" /product= "TDP-4-keto-6-
FT /partial
FT /note= "encodes AAB82201"
FT /*tag= b
FT /gene= "megDVI"
FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT /note= "encodes AAB82202"
FT /*tag= c
FT /gene= "megDII"
FT /product= "TDP-megosamine glycosyltransferase"
FT /note= "eryCIII homologue; encodes AAB80203"
FT /*tag= d
FT /gene= "megY"
FT /product= "mycarose O-acyltransferase"
FT /note= "encodes AAB82204"
FT /*tag= e
FT /gene= "megDII"
FT /product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
FT /note= "eryCI, DnrU homologue, encodes AAB82205"
FT /*tag= f
FT /gene= "megDIII"
FT /product= "daunosaminyl-N'-N'-dimethyltransferase"
FT /note= "eryCVI homologue; encodes AAB82206"
FT /*tag= g
FT /gene= "megDIV"
FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT /note= "eryCVI homologue; encodes AAB82207"
FT /*tag= h
FT /gene= "megDV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
FT /*tag= i
FT /gene= "megBII-1(megDVII)"
FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT /note= "encodes AAB82209"
FT /*tag= j
FT /gene= "megBV"
FT /product= "TDP-mycarose glycosyltransferase"
FT /note= "encodes AAB82210"
FT /*tag= k
FT /gene= "megBIV"
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "encodes AAB82211"
FT /*tag= l
FT /gene= "megA"
FT /product= "megalomiclin 6-deoxyerythronolide B synthase 1"
FT /note= "polyketide synthase; encodes AAB82212"
FT /*tag= m
FT /gene= "megA"
FT /function= "AT-L"
FT /*tag= n
FT /gene= "megA"
FT /function= "ACP-L"
FT /*tag= o
FT misc_feature
FT misc_feature
FT misc_feature
```



```
Db 37866 GCGTCGATGCCGAGCGCAAGCTCACGAGCATGTCCGCCGCTGTCCGAGCGAGGTG 37925
Qy 421 GCACGGTGTCTGGGCGCCAGCGGACCGGATGTCAGAGGTGACGCTTCGCCGAC 480
Db 37926 GCAGCGGTGTCTGGGCGCCAGCGGACCGGATGTCAGAGGTGACGCTTCGCCGAC 37985
Qy 481 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTTGCGGAAACCGGCTCGCGCGGTGACCGGG 540
Db 37986 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTTGCGGAAACCGGCTCGCGCGGTGACCGGG 38045
Qy 541 GTCCGGGTGGCGACGACCATCTGTTCGACCAACCGGACAGTGAACCGGCTTCACCGCGAC 600
Db 38046 GTCCGGGTGGCGACGACCATCTGTTCGACCAACCGGACAGTGAACCGGCTTCACCGCGAC 38105
Qy 601 T 601
Db 38106 T 38106

RESULT 3
AAQ46806
ID AAQ46806 standard; DNA; 29879 BP.
XX
AC AAQ46806;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-DEC-1993 (first entry)
XX
DE eryA region of S. erythraea chromosome.
XX
KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
KM erythromycin; condensation; elongation; acyl chain growth;
KM gene replacement; ss.
XX
OS Saccharopolyspora erythraea.
XX
FH Key Location/Qualifiers
FT CDS 3..10199
FT FT /*tag= a
FT FT /note= "ORF 1"
FT CDS 10218..20921
FT FT /*tag= b
FT FT /note= "ORF 2"
FT CDS 20922..29879
FT FT /*tag= c
FT FT /note= "ORF 3"
XX
PN MO9313663-A1.
XX
PD 22-JUL-1993.
XX
PF 17-JAN-1992; 92WO-US000427.
XX
PR 17-JAN-1992; 92WO-US000427.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Katz L, Donadio S, Mcalpine JB;
XX
DR MPI: 1993-242804/30.
DR P-PSDB; AAR44430, AAR44431, AAR44432.
XX
PT Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -
PT by introducing altered biosynthetic gene-contg. DNA into microorganisms.
XX
PS Claim 27; Fig 2; 13pp; English.
XX
CC This sequence represents a fragment of the Saccharopolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region are
CC involved in the biosynthesis of the polyketide segment of erythromycin.
CC eryA is organised in modules and each module takes care of one
CC condensation step. The precise succession of elongation steps is dictated
```

```
CC by the genetic order of the modules. This fragment may be specifically
CC altered such that novel polyketide molecules of desired structure are
CC produced. Three types of alteration may be produced; those inactivating a
CC single function in a module which does not arrest acyl chain growth;
CC those inactivating a single function in a module which does affect chain
CC growth; and those affecting an entire module. The mutations may be
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 29879 BP; 3672 A; 10596 G; 11545 C; 4066 T; 0 U; 0 Other;
Query Match 66.2%; Score 397.8; DB 2; Length 29879;
Best Local Similarity 78.9%; Pred. No. 1.1e-57;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Qy 1 GCCTCTCAGCGGCGGCGAAGCGCTTCCTGACGCGCTTCGCGCGCGTGTGCGCGAGT 60
Db 24699 GCGCTCTCAGCGGCGGCGAAGCGCTTCCTGACGCGCTTCGCGCGCGTGTGCGCGAGT 24758
Qy 61 GGGCTGCGCGTCACTCGATCGCTGCGGTCTGTGGGCGCGGAGAAATGCGCGGTACC 120
Db 24759 GGGCGCGCGGTGACGTTCATCGCTGGGGGCTGTGGGCGGGGAGAAATGCGCGGAGAC 24818
Qy 121 GAGGGCGGCACTACTGCGGAGCGGAGGCGCTGCGCGCATGACCGCGCGGCGGCGATC 180
Db 24819 GAGGGCGGCACTACTGCGGAGCGGAGGCGCTGCGCGCATGACCGCGCGGCGGCGATC 24878
Qy 181 GAGGAGCTGCGGAGCAACCGCTGAGCGGCGGAGACCGGCGGTGTGCGGTGTGACCTGAC 240
Db 24879 GAGGAGCTGCGGAGCAACCGCTGAGCGGCGGAGACCGGCGGTGTGCGGTGTGACCTGAT 24938
Qy 241 CGGAGGCGGTTCGTGCACTGTTTACCGCGCGCGCGCGCGCGCGCTTTCGACGAACTC 300
Db 24939 CGGAGGCGGTTCGTGCACTGTTTACCGCGCGCGCGCGCGCGCGCTTTCGACGAACTC 24998
Qy 301 GGTGGGCTCGCGCGCGCGCGCGGAGAGACCGGTTCAGGAAATGCGATCTTCGCGCGGCTG 360
Db 24999 GCGGCTCGCGCGCGCGGAGGCGCGGAGAGGAGGAGGAGGCGCGGCTTCGCGCGGCTC 25058
Qy 361 GCGTCGATGCCGAGGCGCGAAGCGTCACGAGCATGTGCGCGCGCGCTGTCGAGCGAGGTG 420
Db 25059 GCGGCTCGTCGACGCGCGAGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25118
Qy 421 GCACGGTGTCTGGGCGCCAGCGGACCGGATGTCAGAGGTGACGCTTCGCTGAC 480
Db 25119 GCGCGGTGTCTGGGCGCCAGCGGACCGGATGTCAGAGGTGACGCTTCGCTGAC 25178
Qy 481 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTTGCGGAAACCGGCTCGCGCGGTGACCGGG 540
Db 25179 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTTGCGGAAACCGGCTCGCGCGGTGACCGGG 25238
Qy 541 GTCCGGGTGGCGACGACCATCTGTTCGACCAACCGGACAGTGAACCGGCTTCACCGCGAC 600
Db 25239 GTCCGGGTGGCGACGACCATCTGTTCGACCAACCGGACAGTGAACCGGCTTCACCGCGAC 25298
Qy 601 T 601
Db 25299 T 25299

RESULT 4
ADM97245
ID ADM97245 standard; DNA; 9510 BP.
XX
AC ADM97245;
XX
DT 01-JUL-2004 (first entry)
XX
DE Synthetic polyketide synthase gene method related synthetic DB853 DNA.
XX ds; gene; polyketide synthase; PCR; medicine; agriculture;
KW synthetic gene; DB853.
XX
```

OS Synthetic.
 XX WO2004029220-A2.
 XX 08-APR-2004.
 XX 26-SEP-2003; 2003WO-US030940.
 XX 26-SEP-2002; 2002US-0414085P.
 XX (KOSA-) KOSAN BIOSCIENCES INC.
 PA Santi DV, Reid RC, Kodumal SJ, Jayaraj S;
 PI WPI; 2004-316102/29.
 XX
 XX
 PT New synthetic genes encoding polyketide synthases, useful for
 PT facilitating production of therapeutic polyketide compounds, or in human
 PT and veterinary medicine, pharmacology, agriculture and molecular biology.
 XX
 XX Example 7; Page 110-112; 172pp; English.
 XX
 CC The present invention relates to a synthetic gene encoding a polypeptide
 CC segment that corresponds to a reference polypeptide segment encoded by a
 CC naturally occurring gene. In particular, the gene is a polyketide
 CC synthase coding sequence. The composition and methods are useful for
 CC producing synthetic genes or libraries of such genes, and for
 CC manipulating and characterizing genes and corresponding encoded
 CC polypeptides. The synthetic genes may be used for facilitating production
 CC of therapeutic or commercially important polyketide compounds. These may
 CC be used in human and veterinary medicine, pharmacology, agriculture and
 CC molecular biology. The present sequence is a synthetic gene used in the
 CC exemplification of the invention.
 XX
 XX Sequence 9510 BP; 1587 A; 2635 C; 3236 G; 2052 T; 0 U; 0 Other;

Query Match 46.8%; Score 281; DB 12; Length 9510;
 Best Local Similarity 66.7%; Pred. No. 3.2e-38;
 Matches 401; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 1 GCCTCTACGCGGCGGGAACGCTTCTCGACGCTTCCGCTGCTGCTGCGCGCAGT 60
 DB 3778 GCCTCTACGCGGCGGGAACGCTTCTCGACGCTTCCGCTGCTGCTGCGCGCAGT 3837
 QY 61 GGGCTGCGCGTCACTCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 DB 3838 GGGCTGCGCGTCACTCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3897
 QY 121 GAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 DB 3898 GAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3957
 QY 181 GAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 DB 3958 GAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4017
 QY 241 CGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 DB 4018 CGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4077
 QY 301 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 DB 4078 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4137
 QY 361 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 4138 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4197
 QY 421 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 DB 4198 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4257
 QY 481 GGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

DB 4258 CTGGGGTTTGAATCTGATGCTGCGGGAACGCTGCTGCGAGCGGCGG 4317
 QY 541 GTCCGGGTGCGCACACGATGCTTTCGACACCCGAGTGAACCGCTTCACCGCGAC 600
 DB 4318 GTACGTGAGGCTGCGCACAGTGTATTGACCATTCACGATCACCGCTTGGCGGATCAT 4377
 QY 601 T 601
 DB 4378 T 4378

RESULT 5
 AAD55818
 ID AAD55818 standard; DNA; 4725 BP.
 XX
 AC AAD55818;
 XX
 DT 27-OCT-2003 (revised)
 DT 07-AUG-2003 (first entry)
 XX
 DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #8.
 XX
 KM Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;
 KM gene; ds.
 XX
 OS Micromonospora carbonacea.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..4725
 FT /tag= a
 FT /product= "Polyketide synthase"
 XX
 XX CA2391131-A1.
 XX
 PD 19-NOV-2002.
 XX
 PE 26-JUL-2002; 2002CA-02391131.
 PR 26-JUL-2001; 2001US-0307629P.
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 PI Vang X, Steffa A, Parnet CM;
 DR WPI; 2003-343556/33.
 DR P-PSDB; AAE37002.
 XX

PT Novel isolated polypeptide involved in biosynthesis of macrolides by
 PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
 PT preferably for biosynthesis of rosamycin.
 XX
 PS Claim 1; Page 164-166; 206pp; English.
 XX

CC The invention relates to genes and proteins involved in the biosynthesis
 CC of macrolides by microorganisms. In particular it relates to the nucleic
 CC acids forming the biosynthetic locus for rosamycin (a 16-member
 CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
 CC useful for the biosynthesis of macrolides by microorganisms. It allows
 CC direct manipulation of macrolides and related chemical structures by
 CC chemical engineering of the proteins involved in the biosynthesis of
 CC rosamycin. It is useful to catalyze certain biochemical reactions, in
 CC vitro or in vivo, to direct or enhance the synthesis or modification of a
 CC polyketide, polyketide substrate or its precursor. The present sequence
 CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
 CC OCT-2003 to standardise OS field)
 XX

SO Sequence 4725 BP; 482 A; 1782 C; 1885 G; 576 T; 0 U; 0 Other;
 Query Match 36.6%; Score 220.2; DB 10; Length 4725;
 Best Local Similarity 61.6%; Pred. No. 4.5e-28;
 Matches 370; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY	3	CTCCAGACGGGCGGGCAACGGCTTCTGTGAAGCCCTGCGCCCGTGTGGTCCGCGCAATGG	62
Db	3879	CGCTTACGGGGGGCGCAACGGCGCCCTTGACGGCTGTCCGACAGCGCGCGCGCCGCCG	3938
QY	63	GCTGCGCGGTCAACTTCGATCGCTGGGGTCTGTGGGCGGGCAGAACATGCGCGTACCGA	122
Db	3999	GCTGACCGCGCACTCTCGTGGCGCGCTGTGGGGCTCTGGGGGGCAATGGCGAGGGCGA	3998
QY	123	GCGCGCGCACTAAGCTTGGCGCAGCGGCTTGGCGCCATGGACCCGACGCGGGCGATGGA	182
Db	3999	CGCGCAGAGTAACTTGACCGCGCGGGGCTTGGCGCGATGCTCTCCGCGCGTGGCGTGGGA	4058
QY	183	GGAAGTGGGAGCAACCTTGGAGCGCGGGGGAACCGTGGGTGTGGGTGGAGCTTGGACCG	242
Db	4059	CGCCCTCTTGGGCGCCCTTGGAGACGGGAGCAAGACTTGTGTGGCGGTGCGCAGTGGACTG	4118
QY	243	GGAAGCGTTTGTGTCGAACTGTTTCAACCGCGCGCCCGCGCGCGCCCTTTCGACGAATCGG	302
Db	4119	GACGCTGTTTCAACGCGCGGGGTTTCAACCGCGTTTCGGCGCCAGCCCGCTGCTCGCGCACTGCC	4178
QY	303	TGGGGTCCGCGCGGGGGCCGAGAGACCGGTC--AGGAATGGATCTTGCCCGGGCGGT	359
Db	4179	GGAAGCCCGCGCGACGCTGGCGGACCGCGGACCCGCGGGCTTCGACCTGCGCGGCTGGCA	4238
QY	360	GGCGTTCGATGCCGAGAGCGCGAAGCTTCAAGAGATGTGCGCGCGGTGTCCGAGCCGAGGT	419
Db	4239	CGCGCCCGGAGCCCCCGAAGAAAGCGCCCGGGGCTGTCTGCACTGTGTACGCGCGCAAGT	4298
QY	420	GGCAGCGGTGTGGGCGCAACCGGACCGCGACGGTGTATCGAGCGTGAAGCTTTCCTTGA	479
Db	4239	CGCGCCCGTCTCTCGGCGCAACCGGGGGCCGAGAGAGTGGGCCCGACGCGCGGTTTCGGGA	4358
QY	480	CCTGGGATTCGACTTCATGACCGCGCGTGGACTCTGCGAACCGGCTCGCGGCGTGAACCGG	539
Db	4359	GATCGGATTCGACTGCTGACCGCGCGTGGACTCTGCGCAAGCGGCTTCAGGGCGGCGTTCGG	4418
QY	540	GGTCCGGGTGGCGCAAGCAATGTCTTTCGACCAACCGGACAGTGGACCGCTTCACCGCGCA	599
Db	4419	CGTGGCGGTGTCCGCAACCTGTGTTCGACCAACCGGACCGGAGCGGCGTTCGCGAGCA	4478
QY	600	C 600	
Db	4479	C 4479	
RESULT 6			
AAD55810			
ID	AAD55810 strand; DNA; 60196 BP.		
XX	AAD55810;		
AC			
XX	27-OCT-2003 (revised)		
DT	07-AUG-2003 (first entry)		
DT			
XX	Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.		
DB			
XX	Macrolide; rosaramycin; polyketide; polyketide synthase; PKS; enzyme;		
KM	gene; de.		
XX			
OS	Micromonospora carbonacea.		
XX			
XX			
Key	Location/Qualifiers		
FT	1..1683		
FT	/*tag= a		
FT	/product= "polyketide synthase #1"		
FT	/note= "CDS does not include start codon"		
FT	/partial		
FT	complement(1728..2522)		
FT	/*tag= b		
FT	/product= "polyketide synthase #2"		
FT	complement(2629..3861)		
FT	/*tag= c		
FT	/product= "polyketide synthase #3"		
FT	CDS		

FT	CDS	4365. .5573
FT		/*tag= d
FT		/product= "Polyketide synthase #4"
FT	CDS	5702. .19117
FT		/*tag= e
FT		/product= "Polyketide synthase #5"
FT	CDS	19144. .24921
FT		/*tag= f
FT		/product= "Polyketide synthase #6"
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	24993. .36230
FT		/*tag= g
FT		/product= "Polyketide synthase #7"
FT	CDS	36292. .41016
FT		/*tag= h
FT		/product= "Polyketide synthase #8"
FT	CDS	41049. .46403
FT		/*tag= i
FT		/product= "Polyketide synthase #9"
FT	CDS	46400. .47794
FT		/*tag= j
FT		/product= "Polyketide synthase #10"
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	47794. .49083
FT		/*tag= k
FT		/product= "Polyketide synthase #11"
FT	CDS	49092. .49814
FT		/*tag= l
FT		/product= "Polyketide synthase #12"
FT	CDS	49868. .51226
FT		/*tag= m
FT		/product= "Polyketide synthase #13"
FT		/transl_except= (pos:51221. .51223, aa:Xaa)
FT		/note= "Xaa corresponds to amino acids from
FT		1811"
FT	CDS	51506. .53416
FT		/*tag= n
FT		/product= "Polyketide synthase #14"
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(54569. .53358)
FT		/*tag= o
FT		/product= "Polyketide synthase #15"
FT	CDS	54897. .56342
FT		/*tag= p
FT		/product= "Polyketide synthase #16"
FT	CDS	56408. .57634
FT		/*tag= q
FT		/product= "Polyketide synthase #17"
FT	CDS	57657. .59123
FT		/*tag= r
FT		/product= "Polyketide synthase #18"
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	59363. .60196
FT		/*tag= s
FT		/product= "Polyketide synthase #19"
XX		
PN		CA2391131-A1.
XX		
PD		19-NOV-2002.
XX		
PF		26-JUL-2002; 2002CA-02391131.
XX		
PR		26-JUL-2001; 2001US-0307629P.
XX		
PA		(ECOP-) ECOPIA BIOSCIENCES INC.
XX		
PI		Yang X, Staefa A, Farnet CM;
XX		
DR		WPI; 2003-343556/33.

Db 4294 TGAAGAGCGCTTCGCTCCGCTTACCTCCGCGCGCCGACAGGCCGCTGTCGACAGACTTG 4353
Qy 301 GGTGGGGTCCG-----CGCCGGGGCGGAGAGACCGGTTCAG 336
Db 4354 CCCGAGGCGCGGAGGTCAATGACGCCACCGCGACGAGGCGGAGAGACACCGGACGC 4413
Qy 337 GAATCGATCTCGCCCGCGGCTGAGCGTCAATGCGGAGGCGGAACCTCAAGCATATGC 396
Db 4414 GCGCCCGCGCTGCGCCGACGACTACCGCGCGCGCCGAGGCGGAACGAGCCACTGCTC 4473
Qy 397 GCCCGCTGTGTCGAGCGGAGGTGACAGCGGTGCTGGGCGACGCGCACCGCATATC 456
Db 4474 CTGCAACTGTGTCGACCGCGCGCTGCGCGCGCTCTCCGCTACGCGGCGCCGACGCGATC 4533
Qy 457 GAGGTGACGTCGCTTCCTTCGACCTGAGATTCGACTTCACGACCGCGCTGACCTGCG 516
Db 4534 GAGCGGCGCGCGGCTTCACAGAGCTGAGCTTCGACTCCCTCACCTCCGTGAACTGCGC 4593
Qy 517 AACCGGCTCGCGCGGAGTACCGGAGGCTCGGAGTGGCCAGACCATGCTTTCACACCG 576
Db 4594 AACCGCTGAAACGCGCGCGGCGGCTCAAGCTGCGCGCCACCTCTGCTTCGACACCGC 4653
Qy 577 ACAGTGACCGGCTTCACCGCGCAC 600
Db 4654 ACGCCCAACCGTCTCGCGCGCAC 4677

RESULT 8
AEB87003
ID AEB87003 standard; DNA; 9684 BP.

AC AEB87003;
AT 06-OCT-2005 (first entry)

XX Streptomyces albusensis polyketide synthase ORF16 DNA SEQ ID NO:34.
XX gene; ds.
XX Streptomyces albusensis.
XX OS
XX ST Streptomyces albusensis.
XX Key Location/Qualifiers
XX FT 1..9684
XX CDS /tag=a
XX FT /product= "polyene polyketide synthase ORF16"

CA2453071-A1.
PD 03-APR-2004.
XX 21-JAN-2004; 2004CA-02453071.
XX 21-MAY-2003; 2003US-0441123P.
XX 13-MAY-2003; 2003US-0469810P.
XX 01-AUG-2003; 2003US-0491516P.
XX 13-AUG-2003; 2003US-0494568P.
XX (ECOP-1) ECOP1A BIOSCIENCES INC.
XX Farnet CM, Zazopoulos E;
XX WPI; 2004-348649/33.
XX P-PSDB; AEB87002.
XX Isolated, purified or enriched nucleic acid useful for producing
XX polyketides, encodes domains of the polyketide synthase system.
XX Claim 2; SEQ ID NO 34; 550bp; English.
XX The invention relates to an isolated, purified or enriched nucleic acid
XX which can be used for the production of a polyketide. The nucleic acid

CC encodes at least one domain of the polyketide synthase system formed by
CC the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,
CC AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)
CC an isolated, purified or enriched nucleic acid for the production of a
CC polyketide of Formula I comprising a nucleic acid selected from the group
CC consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,
CC AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched
CC expression vector comprising the novel nucleic acid or the nucleic acid
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC the cosmid deposited under INAC accession no. 250203-01, 250203-02,
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC the production of a polyketide. The nucleic acid can be used for
CC production of a polyketide, for use as a fungal and cancer cell growth
CC inhibitors. The present sequence encodes Streptomyces albusensis polyene
CC polyketide synthase ORF16, which is used in the exemplification of the
CC present invention.

XX Sequence 9684 BP; 1211 A; 3372 C; 3596 G; 1505 T; 0 U; 0 Other;
SQ

Query Match 34.5%; Score 207.6; DB 13; Length 9684;
Best Local Similarity 60.3%; Pred. No. 5,2e-26;
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

Qy 1 GCTCTTACGCGCGGCGCAACGCTTCTTCAAGCGCTTTCGCGCTGCTGTCGCGCACT 60
Db 4054 GCGCACTACGCGCGGCGCAACGCTTCTTCAAGCGCTTTCGCGCTGCTGTCGCGCACT 4113
Qy 61 GGGTGGCGGCTCACTGAGTACGCTGAGGCTGAGGCGCGGCGGCGGCGGCGGCGGCGG 120
Db 4114 GAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4173
Qy 121 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 4174 GCGGTGCGTGAACCGGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4233
Qy 181 GAGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 4234 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4293
Qy 241 GCGGAGCGGCTTCTGAGTACGCTGAGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 4294 TGAAGAGCGCTTCGCTTCGCTTCACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4353
Qy 301 GGTGGGGTCCG-----CGCCGGGGCGGAGAGACCGGTTCAG 336
Db 4354 CCCGAGGCGCGGAGGTCAATGACGCCACCGCGACGAGGCGGAGAGACACCGGACGC 4413
Qy 337 GAATCGATCTCGCCCGCGGCTGAGCGTCAATGCGGAGGCGGAACCTCAAGCATATGC 396
Db 4414 GCGCCCGCGCTGCGCCGACGACTACCGCGCGCGCCGAGGCGGAACGAGCCACTGCTC 4473
Qy 397 GCCCGCTGTGTCGAGCGGAGGTGACAGCGGTGCTGGGCGACGCGCACCGCATATC 456
Db 4474 CTGCAACTGTGTCGACCGCGCGCTGCGCGCGCTCTCCGCTACGCGGCGCCGACGCGATC 4533
Qy 457 GAGGTGACGTCGCTTCCTTCGACCTGAGATTCGACTTCACGACCGCGCTGACCTGCG 516
Db 4534 GAGCGGCGCGGCTTCACAGAGCTGAGCTTCGACTCCCTCACCTCCGTGAACTGCGC 4593
Qy 517 AACCGGCTCGCGCGGAGTACCGGAGGCTCGGAGTGGCCAGACCATGCTTTCACACCG 576
Db 4594 AACCGCTGAAACGCGCGCGGCGGCTCAAGCTGCGCGCCACCTCTGCTTCGACACCGC 4653
Qy 577 ACAGTGACCGGCTTCACCGCGCAC 600
Db 4654 ACGCCCAACCGTCTCGCGCGCAC 4677

RESULT 9
AEB86800
ID AEB86800 standard; DNA; 9684 BP.

FT	/product= "Polyketide synthase ORF15 protein"
FT	94057. .103740
FT	/*tag= h
FT	/product= "Polyketide synthase ORF16 protein"
FT	103789. .126321
FT	/*tag= i
FT	/product= "Polyketide synthase ORF17 protein"
FT	126389. .138007
FT	/*tag= j
FT	/product= "Polyketide synthase ORF18 protein"
FT	complement(138063. .139079)
FT	/*tag= k
FT	/product= "Polyketide synthase related ORF19 protein"
FT	complement(139266. .140117)
FT	/*tag= l
FT	/product= "Polyketide synthase related ORF20 protein"
FT	complement(140114. .141103)
FT	/*tag= m
FT	/product= "Polyketide synthase related ORF21 protein"
FT	141483. .142436
FT	/*tag= n
FT	/product= "Polyketide synthase related ORF22 protein"
FT	142440. .143054
FT	/*tag= o
FT	/product= "Polyketide synthase related ORF23 protein"
FT	143133. .144119
FT	/*tag= p
FT	/product= "Polyketide synthase related ORF24 protein"
FT	144116. .145102
FT	/*tag= q
FT	/product= "Polyketide synthase related ORF25 protein"
FT	145099. .145743
FT	/*tag= r
FT	/product= "Polyketide synthase related ORF26 protein"
FT	145818. .147230
FT	/*tag= s
FT	/product= "Polyketide synthase related ORF27 protein"
FT	complement(147306. .148967)
FT	/*tag= t
FT	/product= "Polyketide synthase related ORF28 protein"
FT	complement(149176. .149871)
FT	/*tag= u
FT	/product= "Polyketide synthase related ORF29 protein"
FT	complement(149868. .150789)
FT	/*tag= v
FT	/product= "Polyketide synthase related ORF30 protein"
FT	complement(150769. .153765)
FT	/*tag= w
FT	/product= "Polyketide synthase related ORF31 protein"
FT	154485. .156041
FT	/*tag= x
FT	/product= "Polyketide synthase related ORF32 protein"
FT	156075. .157064
FT	/*tag= y
FT	/product= "Polyketide synthase related ORF33 protein"
FT	157308. .158873
FT	/*tag= z
FT	/product= "Polyketide synthase related ORF34 protein"
FT	158970. .160202
FT	/*tag= aa
FT	/product= "Polyketide synthase related ORF35 protein"
FT	160199. .161719
FT	/*tag= ab
FT	/product= "Polyketide synthase related ORF36 protein"
FT	161924. .162577
FT	/*tag= ac
FT	/product= "Polyketide synthase related ORF37 protein"
FT	162723. .164051
FT	/*tag= ad
FT	/product= "Polyketide synthase related ORF38 protein"
XX	
PN	WO2004065401-A1.

[illegible]

PI Farnet CM, Bachmann BO, Mcalpine JB, Zazopoulos E;
XX WPI; 2004-348650/33.
XX
XX Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT for treating e.g. fungal infections and cancer.
XX
XX Disclosure; SEQ ID NO 18; 144pp; English.
XX
XX The invention relates to an isolated, purified or enriched nucleic acid
CC which can be used for the production of a polyketide. The nucleic acid
CC encodes at least one domain of the polyketide synthase system formed by
CC the polyketide synthases of ABB86662, ABB86664, ABB86666, ABB86668,
CC ABB86670, ABB86672, ABB86674, ABB86676 and ABB86678. Also described: (1)
CC an isolated, purified or enriched nucleic acid for the production of a
CC polypeptide of formula I comprising a nucleic acid selected from the group
CC consisting of ABB86663, ABB86665, ABB86667, ABB86669, ABB86671, ABB86673,
CC ABB86675, ABB86677 and ABB86679; (2) an isolated, purified or enriched
CC nucleic acid for the production of a polyene polyketide; (3) an
CC expression vector comprising the novel nucleic acid or the nucleic acid
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC the comid deposited under IDMC accession no. 250203-01, 250203-02,
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC the production of a polyketide. The nucleic acid can be used for
CC production of a polyketide, for use as a fungal and cancer cell growth
CC inhibitors. The present sequence represents a Streptomyces albidus
CC polyene polyketide synthase related contig 2 DNA sequence, which is used
CC in the exemplification of the present invention.
XX
XX Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;
SQ
Query Match 34.5%; Score 207.6; DB 13; Length 164051;
Best Local Similarity 60.3%; Pred. No. 3.7e-26;
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;
QY 1 GCCTCTACGCGCGCGGACGCTTCTCGAGCGCTTCCGCGCTGCTGCGCGAGT 60
DB 98110 GGCACACTACGCGCGCGGACGCGCTTCTCGAGCGCTTCCGCGAGTACCGCGCGAGTCC 98169
QY 61 GCGCTCGCGGATCACTCGATCGCTGCGGCTGCTGCGCGCGGACGAGATGCGCGTAC 120
DB 98170 GGACTGCGCGGACCTCGATCGCTGCGGCGCTTCTGCGCGGACGAGGAGTCCGCGAGG 98229
QY 121 GAGGCGCGGATCACTCGATCGCTGCGGCGCTGCGGCGGACGAGGAGTCCGCGAGTCC 180
DB 98230 GCGGCTCGGATCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98289
QY 181 GAGGAGCTGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 98290 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98349
QY 241 CGGAGCGGATGCTGCGGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 98350 TCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 98409
QY 301 GGTGGGCTCGG-----CGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 336
DB 98410 CCGGAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98469
QY 337 GAATCGGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
DB 98470 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98529
QY 397 GCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
DB 98530 CTCGAATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98589
QY 457 GAGCGTGAAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 516
DB 98590 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98649
QY 517 AACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576

DB 98650 AACCGGCTGAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98709
QY 577 ACAGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 98710 ACGCCACCGCTCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 98733
RESULT 13
AAD55819
ID AAD55819 standard; DNA; 5355 BP.
XX
XX AAD55819;
AC
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DB Micromonospora carbonacea polyketide synthase (PKS) type I gene #9.
XX
XX Macrolide; roxamycin; polyketide; polyketide synthase; PKS; enzyme;
KM gene; ds.
XX
OS Micromonospora carbonacea.
XX
XX Key Location/Qualifiers
FH CDS 1..5355
FT
FT /product= "Polyketide synthase"
XX
XX CA2391131-A1.
XX
XX 19-NOV-2002.
XX
XX 26-JUL-2002; 2002CA-02391131.
XX
XX 26-JUL-2001; 2001US-0307629P.
XX
XX (BCOP-) ECOMA BIOSCIENCES INC.
XX
XX Yang X, Staffa A, Farnet CM;
PI WPI; 2003-343556/33.
DR P-PSDB; AAE37003.
XX
XX Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of roxamycin.
XX
XX Claim 1; Page 176-176; 206pp; English.
XX
XX The invention relates to genes and proteins involved in the biosynthesis
CC of macrolides by microorganisms. In particular it relates to the nucleic
CC acids forming the biosynthetic locus for roxamycin (a 16-member
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC useful for the biosynthesis of macrolides by microorganisms. It allows
CC direct manipulation of macrolides and related chemical structures by
CC chemical engineering of the proteins involved in the biosynthesis of
CC roxamycin. It is useful to catalyze certain biochemical reactions, in
CC vitro or in vivo, to direct or enhance the synthesis or modification of a
CC polyketide, polyketide substrate or its precursor. The present sequence
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
CC OCT-2003 to standardise OS field)
XX
SQ Sequence 5355 BP; 516 A; 2056 C; 2107 G; 676 T; 0 U; 0 Other;
Query Match 34.0%; Score 204.4; DB 10; Length 5355;
Best Local Similarity 60.0%; Pred. No. 1.9e-25;
Matches 359; Conservative 0; Mismatches 236; Indels 3; Gaps 1;
QY 3 CTCCTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62
DB 3876 CGGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9395
QY 63 GCTGCGGCTACCTGATCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122

Db 3936 GCTGCCCGGACCTCCGCTCGCTGGGGGCTGGGGGCGGGGCGATGGGGGCGGCGC 3995
Qy 123 GGGCGGGGACCTACTGGGGGACCGGCGCTGCGGCGCATGAGCCCGGCGGCGATCGA 182
Db 3996 GGGGGAGAGATCTGGGCGCGCGCGGGCGCTCCGGGCGCATGCGCCCGGCGAGCGCCCTCGC 4055
Qy 183 GGAAGCTCGGAGCAACCTTGAACCGCGGGGAGCCCGTGGGTGTGGGTGAGCCTGGACCG 242
Db 4056 CGCCCTGGGGCGGATCTTGAACCGCGAGAGACCGGGGTGACGGGTCTCCGACCGAGTG 4115
Qy 243 GGAAGCGGTTCGTGAACTGTTCACCGCGCGCGCGCGCGCCCTCTTTCGACGAACTCGG 302
Db 4116 GGGCGGGTTCGCGCTCGCGCTTCCGCGCGCGCTCCGCGCGCTGCTCGCGAGCTGCC 4175
Qy 303 TGGGGTTCGCGCGCGGGGCGGAGAGACCGGTCAAGAAATCGATTCGCGCGGGCTGGC 362
Db 4176 GGGCGGGGAGATCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4235
Qy 363 GTCGATGCGGAGGCGGAGCGTCAACGAGCAATGTGCGCGGTGGTCCGAGCGGAGTGCG 422
Db 4236 GGGCGGGGCGGAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4292
Qy 423 AGCGGTGCTGGGCGGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 482
Db 4293 CGGGGTGCTGGGCGGAGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4352
Qy 483 GGAATTCGATCCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 542
Db 4353 CGGCTTCGATCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4412
Qy 543 CGGGGTGCGGACGACATGCTTTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 600
Db 4413 GAGCTCTCGGCGCGCGCGCTGCTGTCGACGCGCGCGCGCGCGCGCGCGCGCGC 4470

RESULT 14
ADY80053
ID ADY80053 standard, DNA; 67167 BP.
AC ADY80053;
XX
XX
XX 02-JUN-2005 (first entry)
DE Nucleotide sequence of the Leptomycin PKS gene cluster.
KW Leptomycin; polyketide synthase; PKS; LepA; LepB; LepC; LepD; LepE; LepF;
KW Leptomycin PKS; polyketide; enzyme purification; ss.
XX
OS Streptomyces sp.
XX
XX
PN MO2005026327-A2.
XX
XX 24-MAR-2005.
PD
PF 10-SEP-2004; 2004MO-US029630.
XX
XX 11-SEP-2003; 2003US-0502423P.
PR 15-MAR-2004; 2004US-055384P.
PR 08-SEP-2004; 52US-00553384.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
PA
XX Hu Z, Reid RC;
PI
XX
XX WPI; 2005-233487/24.
DR
XX
XX
XX New isolated, purified, or recombinant DNA molecule comprising a sequence
PT encoding a domain or activity of an enzyme required for the biosynthesis
PT of Leptomycin, useful for producing polyketides and Leptomycin.
XX
XX Disclosure; SEQ ID NO 3; 112bp; English.
XX

CC The specification describes Leptomycin polyketide synthase (PKS)
CC polynucleotides and polypeptides. These enzymes are required for the
CC biosynthesis of Leptomycin, and are selected from LepA, LepB, LepC, LepD,
CC LepE and LepF. The Leptomycin PKS polynucleotides are useful for
CC producing polyketides, Leptomycin and Leptomycin analogs and derivatives.
XX The present sequence represents a Leptomycin PKS gene cluster.
SQ Sequence 67167 BP; 7789 A; 22625 C; 26867 G; 9866 T; 0 U; 0 Other;
Query Match 33.5%; Score 201.6; DB 14; Length 67167;
Best Local Similarity 59.6%; Pred. No. 4.1e-25;
Matches 361; Conservative 0; Mismatches 239; Indels 6; Gaps 1;
Qy 1 GCCTCTTCAAGCGCGGCGGCAACGCTTCTCTGAGCGCTTCGCGGTGTGGCGGCACT 60
Db 62632 GGGCGCTTACCGCGGCGGCGGCAACGCTTCTTGAAGCGCTTCGCGGCGGCAACGCG 62691
Qy 61 GGGCTGCGGCTTCACTCGATCGCTGGGGGTGTGGGGCGGGGAGAACATGGCGGTACG 120
Db 62692 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62751
Qy 121 GAGGGGCGGCGACTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 62752 GAGGGGAGAACGCGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62811
Qy 181 GAGGAGCTGCGGAGCAACCTTGAAGCGCGGGGAGCCGCGGTGTGGGTGAGACTGAGC 240
Db 62812 GCGCGACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62871
Qy 241 CGGAGGCGGCTGTGCGAATGTTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 299
Db 62872 CGGCGCGGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62931
Qy 300 -----CGTGGGTGCTCGCGCGGCGGCGGCGGCGGAGAGACCGGTTCAGAGATCTGCGCGG 354
Db 62932 ATGAGCGGAGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62991
Qy 355 CGGCTGCGGCTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 414
Db 62992 CGGCGGCTGACCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63051
Qy 415 GAGGTGAGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 474
Db 63052 GAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63111
Qy 475 CGTGAACCTGGAATTCGATCTCAATGACCGCGGCTGCACTGCGGAGACCGGCTCGCGGCGG 534
Db 63112 CGGAGCTTGGGCTTGAATCTGCGTGGCGCGCGGCTGCACTGCGGAGACCGGATCGCGGAGCGC 63171
Qy 535 ACCGGGGTTCGGGGTGGCGGAGCAACATGCTTTCGACCAACCGGAGTGGAGCGGCTTCAAC 594
Db 63172 ACCGGGGTACACTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63231
Qy 595 GCGGCGC 600
Db 63232 GCAAC 63237

RESULT 15
ADX56092
ID ADX56092 standard, DNA; 104096 BP.
AC ADX56092;
XX
XX
XX 21-APR-2005 (first entry)
DE Streptomyces nanchangensis nanoliomycin gene cluster SeqD1.
KW nanoliomycin; agriculture; pharmaceutical; protein production; gene; ds.
XX
XX Streptomyces nanchangensis.
OS
XX
XX Key Location/Qualifiers
FH

Db 51227 GCGCCCGCGGTGATCTGCCGACGTACGCCCTTCCAGCGCGAGCGCTACTGGCTGACG 512

QY	121	TCGCCCTGGGCGGGGTCCCCGACACCGGCGGACGATGCGGTACAGCTCGCCTGGCACC	180
Db	51287	CGCCACCGCCGAGAGCCCCCAACGAGCGCATTCACGCGGAAATCTGGGCGCGCTGCAGAC	51348
QY	181	CCGTGCACTTCGGGCGGTCTCTCCCTGGCGGAGCGGGTCTGTGTGTGACCGAGCGGCGAG	240
Db	51347	GCGAGGACCTTCGAGTGGCTTGGCGGACAGCTGGCGCTTCACGCGGCGAGCGGCTTGGCGAAG	51408
QY	241	TACCCCGGAGCTGGACGGAAGTGGTCCGACGCGCTTGA	280
Db	51407	TGTGCGCCGGCTCTTCCAGATGGGAGGCGCAAGCGCTGA	51446

```

RESULT 2
US-10-858-730-47
Sequence 47, Application US/10858730
Publication No. US20050255568A1
GENERAL INFORMATION:
  APPLICANT: Bailey, Richard B.
  APPLICANT: Blomquist, Paul
  APPLICANT: Dotcen, Reed
  APPLICANT: Driggers, Edward M.
  APPLICANT: Madden, Kevin T.
  APPLICANT: O'Leary, Jessica
  APPLICANT: O'Toole, George
  APPLICANT: Trueheart, Joshua
  APPLICANT: Walbridge, Michael J.
  APPLICANT: Yorgey, Peter S.
  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
  TITLE OF INVENTION: PRODUCTION
  FILE REFERENCE: 14184-030001
  CURRENT APPLICATION NUMBER: US/10/858,730
  CURRENT FILING DATE: 2004-06-01
  PRIOR APPLICATION NUMBER: US 60/475,000
  PRIOR FILING DATE: 2003-05-30
  PRIOR APPLICATION NUMBER: US 60/551,860
  PRIOR FILING DATE: 2004-03-10
  NUMBER OF SEQ ID NOS: 364
  SOFTWARE: FastSeq for Windows Version 4.0
  SEQ ID NO 47
  LENGTH: 1290
  TYPE: DNA
  ORGANISM: Streptomyces coelicolor
US-10-858-730-47

```

Query Match Similarity	11.9%	Score 71.2	DB 8	Length 1290
Best Local Similarity	51.9%	Pred. No. 8	6e-07	
Matches 160	Conservative 0	Mismatches 148	Indels 0	Gaps 0
QY	221	GATGTGACCGGAGCGGACGATACCCCGGCGTGAAGGACGTGATCCGGACGAGCGCTGGA	280	
Db	60	GATGTGCGGACATCATGAGACGACGCGCGGACCTCGCCGCGGGATCGGGGCCCGGT	119	
QY	281	ACAGCGCGGGGCGACCGTGTGTGTGCAACGCGGACAGTGCAGCGCCCGGATCGGCGCGC	340	
Db	120	GAGAGCTCCGGGCGCTCGCCGTACGCGCGGCCCAAGAGTGCAGGAGGGGATGACCCGCGC	179	
QY	341	ACTGAGCGCCGTCGAGCGGACCGGCCCTGTGCACATGTGGCTCTGACTGCGCTGACGGA	400	
Db	180	CCTCGTCAACACCGAGCGCACCGCGCTGTCAAGCGCGGGGACATGACGTGTGTGCA	239	
QY	401	GAGCGAGTCTGTGCAAGCAACCCAGCGTGGACACCTTCGCGTTGTGTCAAGCGCTCGCGC	460	
Db	240	GGTATCGGGGGGAGATCGAGCCCGCGGACGCTATACACACCGCTTCGCCACGCGCC	299	
QY	461	AGCGGGATCGACGTGCCCTGTGTGTGTGTCACAGGAGCGCGCGCGCGTACCGTGG	520	
Db	300	CTCCGTGTTTCGCCCAACAGGCGCTCATCGCCAGGACGAGCGCGCGCTTCACGCGCGC	359	
QY	521	AGAGGACG 528		
Db	360	CGCGGACG 367		

```

RESULT 3
US-11-052-554A-519/c
: Sequence 519, Application US/11052554A
: Publication No. US20050288866A1
: GENERAL INFORMATION:
: APPLICANT: Sachdeva, et al.
: TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
: TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
: FILE REFERENCE: 30853/40359A
: CURRENT APPLICATION NUMBER: US/11/052,554A
: CURRENT FILING DATE: 2005-02-07
: PRIOR APPLICATION NUMBER: US 60/589,227
: PRIOR FILING DATE: 2004-07-20
: PRIOR APPLICATION NUMBER: IN 173/DEL/2004
: PRIOR FILING DATE: 2004-02-06
: NUMBER OF SEQ ID NOS: 763
: SOFTWARE: Patentin version 3.3
: SEQ ID NO 519
: LENGTH: 5706
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-519

```

Query Match	11.2%	Score 67.2	DB 14	Length 5706
Best Local Similarity	47.0%	Pred. No. 4,66-06		
Matches 207	Conservative 0	Mismatches 233	Indels 0	Gaps 0
Qy	124	CCCTGGGCGCGGTCCTCCGACACCGGCGCAGAGTGGCTTACCACTCGCTTGACACCCCG	183	
Db	4970	CCGCGCGGCGCGGCGGTTGCGCGCCGACCCCGCTTGCGCGCTTGAGCGCGCGGCGAAGCGCG	4911	
Qy	184	TTCGACCTTCGGCGGCTCCTCCTCGGCGCGAGCTCTGTGTGTGAACCGAGACGGCAGTAC	243	
Db	4910	TTGCGCTTCTTGGAGAGGCGCGCTGTGCGCGGCGCGCGCTGCGCGCGCGCGCGCG	4851	
Qy	244	CCCCGGCGCTGGACGAGAGCTGTGTCGCGACCGGCTGTGAACAGCGCGGCGGACCGTGTGT	303	
Db	4850	GCCCCCGCGAATCCGCGGCTTGCCGCGGTTCCGCGCGGTGTGCGCAAGCGCGCGCACGCG	4791	
Qy	304	TGTGACCCGCGCAGTCTGCGCGCGCGGATCGGCGCGCACTGACGCGCGTGCAGCGACCG	363	
Db	4790	GCGTTCGCGCGCGATCCGCGCGGATCCGCGCGACCGCGAATAAGCGTCCCGTCCGCG	4731	
Qy	364	CCCTGTCCACTGTGTGTCTCTGTGCTGCGCTGCGCGAGGCGGCGGTGTGCAGCACCCCA	423	
Db	4730	CCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGCGCGCGGTACCGGCTGTGGCGTGGCT	4671	
Qy	424	GCGTGGACACCCCTGCGGTTGTGTCAGGCGCTCGGCGCAGCGGAGTCGACGCTCCCTGT	483	
Db	4670	AGCCCGGAGCTGTTCAGGCGCGCTTTGCGCGCTTGCGCGGTGGCGCGCGGACCACT	4611	
Qy	484	GCGTGTGACCAAGGAGCGCGCGCGCGGTGACCGTGTGAGAGACGAGCTGATCCGCGCCAG	543	
Db	4610	TGGCGCGCGTCAATGCTGGGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4551	
Qy	544	CCATGTGTCGTGGGCTCGGC	563	
Db	4550	CCGCTGCGCGCGATCCGCGC	4531	

APPLICANT: Eteghalian, Alireza
TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
FILE REFERENCE: 56462007901
CURRENT APPLICATION NUMBER: US/10/517, 939
CURRENT FILING DATE: 2004-12-13
PRIOR APPLICATION NUMBER: PCT/US03/19153
PRIOR FILING DATE: 2003-06-16
PRIOR APPLICATION NUMBER: 60/389,299
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 380
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 359
LENGTH: 2724
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample.
US-10-517-939-359

Query Match 10.5%; Score 63; DB 8; Length 2724;
Best Local Similarity 45.5%; Pred. No. 4.4e-05;

Matches 285; Conservative 0; Mismatches 315; Indels 3; Gaps 1;

QY 14 GGCGTACACCCGGGGGTGAGGTGCACTGGCTACCGCAGTGGGTGAGGAGAGCCCGGT 73
DB 2046 GGCGTACACCCGGGGGTGAGGTGCACTGGCTACCGCAGTGGGTGAGGAGAGCCCGGT 2105
QY 74 CGACCTGCGCGGTCTACCGCTTCCAGCAGAACTTCTGCTCCCGGTCCCTGGGCG 133
DB 2106 CGCGAGAGACATGCGCTTCCAGCAGAACTTCTGCTCCCGGTCCCTGGGCG 2165
QY 134 GGTCCTCCGACACCGGGGAGAGGTGGCTTACAGCTTCTGCTGGCAGCCCGCTGACCTCG 193
DB 2166 GGTCCTCCGACACCGGGGAGAGGTGGCTTACAGCTTCTGCTGGCAGCCCGCTGACCTCG 2225
QY 194 GGCGTCTCTGCTGCGGAGAGGGTCTGCTGCTGAGACCGAGGCGGAGTACCCCGGCTG 253
DB 2226 GGCGTCTCTGCTGCGGAGAGGGTCTGCTGCTGAGACCGAGGCGGAGTACCCCGGCTG 2285
QY 254 GACGAGCGTGTCTGCGGAGAGGGTCTGAGAAAGCGGCGGAGCGCTGTGTTGCAACCGC 313
DB 2286 GACGAGCGTGTCTGCGGAGAGGGTCTGAGAAAGCGGCGGAGCGCTGTGTTGCAACCGC 2345
QY 314 GCAATGCGCGCGCGCGGATGCGCGCGCGCACTGACGCGCG--TGACGCGACCGCGCTGTC 370
DB 2346 CTCCTGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2405
QY 371 CACTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
DB 2406 CCACTCTCGGATCATGACGCGCGCACACCTCTCGGCTGGAAGTGTCTGCGG 2465
QY 431 CACCTCTCGGATCATGACGCGCGCACACCTCTCGGCTGGAAGTGTCTGCGG 490
DB 2466 CACCTCTCGGATCATGACGCGCGCACACCTCTCGGCTGGAAGTGTCTGCGG 2525
QY 491 GACGAGGAGCGCGCGCGCGCTGACCGTGGAGACGAGCTGATCCGCGCGCGCGAGTGT 550
DB 2526 GACGAGGAGCGCGCGCGCGCTGACCGTGGAGAGCGCGCGCGCGCGAGTGT 2585
QY 551 CGGTGGGCTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
DB 2586 CGGTATGAGGCGCGCTGCTGACGCGCGAGAGCGCGAGATCCGG 2628

RESULT 5
US-11-052-554A-519

Sequence 519, Application US/11052554A
Publication No. US2005028866A1
GENERAL INFORMATION:
APPLICANT: Sachdeva, et al.
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

FILE REFERENCE: 30853/40359A
CURRENT APPLICATION NUMBER: US/11/052,554A
CURRENT FILING DATE: 2005-02-07
PRIOR APPLICATION NUMBER: US 60/569,227
PRIOR FILING DATE: 2004-07-20
PRIOR APPLICATION NUMBER: IN 173/DEL/2004
PRIOR FILING DATE: 2004-02-06
NUMBER OF SEQ ID NOS: 763
SOFTWARE: PatentIn version 3.3
SEQ ID NO 519
LENGTH: 5706
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-519

Query Match 10.3%; Score 62; DB 14; Length 5706;
Best Local Similarity 49.0%; Pred. No. 6.3e-05;

Matches 193; Conservative 0; Mismatches 200; Indels 1; Gaps 1;

QY 183 GTGACCTCGGGGCGCTCTCCCTGCGCGAGCGGTCTGTGTGACCCGAGGCGAGTA 242
DB 2066 GTGACCTCGGGGCGCTCTCCCTGCGCGAGCGGTCTGTGTGACCCGAGGCGAGTA 2125
QY 243 CCCCCCGCTGACGAGCGTGTCTCGCGACGCGCTTGAACAAGCGCGGCGACCGTCTG 302
DB 2126 GCGCCCGCGAGAGAG--CGAGCGCGCGCGACAGCGGTGTGGCGGCAACAGGCTCC 2184
QY 303 TTGTGACCGCGAGTGTGCGCGCGCGCGATGTGGCGCGCACTGACGCGCTGACGCGACCC 362
DB 2185 GCGCGCGCGCGCGGTGCAAGCGCGCAAGGCGCGCAAGGCGCGCGCGCGCGCGCG 2244
QY 363 GCGCTGTCACTGTGTCTCTGCTGCGGCTTGTGCGAGGCGGCTGTGTGACAGACCC 422
DB 2245 GACAACTTCAAGGTGTGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2304
QY 423 AGCTTGAACACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482
DB 2305 GCG 2364
QY 483 TGGCTGTGACGAGGAGCG 542
DB 2365 GGTATGTCAGGATGTCG 2424
QY 543 GCCATGT 576
DB 2425 GGTCTGCGCTGCGCTTCTGCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2458

RESULT 6
US-11-096-568A-12512

Sequence 12512, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptide
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
CURRENT FILING DATE: 2005-04-01
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 12512
LENGTH: 1450
TYPE: DNA
ORGANISM: Trifolium aestivum
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(1450)
OTHER INFORMATION: Ceres Seq. ID no. 14302059
US-11-096-568A-12512

Query Match 10.2%; Score 61; DB 11; Length 1450;
Best Local Similarity 45.0%; Pred. No. 0.00014;

Matches 229; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

LOCATON: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

Query Match 10.1%; Score 60.4; DB 14; Length 88421;
Best Local Similarity 49.1%; Pred. No. 7.8e-05;
Matches 160; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

191 CGGCGGATCTCTCTGCGCGGAGCGGATCTGTGTGACCGGAGCGGAGTACCCCGGCG 250
16476 CGGCGTGTGCGGCGCTCCCGCGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 16535

251 CTGAGACGAGCTGTGCGGAGCGGCGGCTGGAACAGCGGCGGCGGAGCGGCTGTGTTGACAC 310
16536 CGAGACGCGGCTGTGCGGAGCGGAGCAACCTCTTCCGCGGAGCGGCGGCGGCGGCGGCGGCG 16595

311 CGGCGAGTGTGCGGCGGCGGAGTGTGCGGCGGAGTGTGAGCGGCGGAGCGGCGGCGGCGGCGG 370
16596 GGGGGGCGGCGGCGGCTGTGCGGAGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 16655

371 CACTGTGTCTCTCTGCTGTGCGGCTGTGCGGAGCGGCGGCTGTGCGGAGCGGCGGCGGCGGAG 430
16656 CCAAGCAACGAGTGTGCGGCGGCGGCGGAGCGGCTGTGCGGAGCGGCGGCGGCGGCGGAGC 16715

431 CACCTTGTGCTGTGCGGAGCGGCTGTGCGGCGGAGCGGCGGAGTGTGCGGCGGCGGCGGCGGCGG 490
16716 CACCTTGTGCTGTGCGGAGCGGCTGTGCGGCGGAGCGGCGGAGTGTGCGGCGGCGGCGGCGGCGG 16775

491 GACCAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 516

16776 GGTCAAGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 16801

US-11-096-568A-8975
Sequence 8975, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
FILE REFERENCE: 2750-15927PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 8975
LENGTH: 1328
TYPE: DNA
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(1328)
OTHER INFORMATION: Ceres Seq. ID no. 13596902
US-11-096-568A-8975

Query Match 10.0%; Score 60.2; DB 11; Length 1328;
Best Local Similarity 47.9%; Pred. No. 0.00021;
Matches 173; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

104 GAACCTGTGAGTCTCCGCGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 163
169 GCACTCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 228

164 CCACTGTGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 223
229 CTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 288

224 GGTGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 283
289 GCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 348

284 GCGCGG 343
349 CCGCGG 408

344 CCAAGCGG 403
409 CAAAGCGG 468

404 CCGTGTGTGTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 463
469 CATCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 528

464 C 464
529 C 529

RESULT 9
US-11-052-554A-543/C
Sequence 543, Application US/11052554A
Publication No. US2005028886A1
GENERAL INFORMATION:
APPLICANT: Sachdeva, et al.
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
FILE REFERENCE: 30853/40359A
CURRENT APPLICATION NUMBER: US/11/052,554A
PRIOR FILING DATE: 2005-02-07
PRIOR APPLICATION NUMBER: US 60/589,227
PRIOR FILING DATE: 2004-07-20
PRIOR APPLICATION NUMBER: IN 173/DEL/2004

;; PRIOR FILING DATE: 2004-02-06
;; NUMBER OF SEQ ID NOS: 763
;; SOFTWARE: Patentin version 3.3
;; SEQ ID NO 543
;; LENGTH: 2514
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-543

Query Match 10.0%; Score 59.8; DB 14; Length 2514;
Best Local Similarity 47.9%; Pred. No. 0.00022;
Matches 172; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 170 CGCCTGACACCCCGGACCTCGGACCGGCTCTCCCTGACCGAAGGATCTGTGTGAC 229
DB 733 CGCCGCGCCCGCGGATCGCGGCTCGGTCGCCCGCGCCACCGACCCCGCGCTCG 674
QY 230 CGAAGCGGCAATACCCCGCGCTGGAACGAAGTGTCCGCGACCGGCTGGAACAGCGCG 289
DB 673 CGAAGCGGCGGCTTTCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 614
QY 290 GCGACCGCTGT 349
DB 613 CGAATCGCGCGGACCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
QY 350 CGTGAACGCGACCGCGCGCTGT 409
DB 553 CGCT 494
QY 410 TGTGACGACCGCGCGCTGT 469
DB 493 TGCAGTAGAGCG 434
QY 470 CGACGTCCCTGT 528
DB 433 CGAATCTCTCG 375

RESULT 10

US-11-052-554A-546
; Sequence 546; Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 546
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-546

Query Match 9.9%; Score 59.4; DB 14; Length 1632;
Best Local Similarity 49.6%; Pred. No. 0.0003;
Matches 180; Conservative 0; Mismatches 181; Indels 2; Gaps 1;

QY 207 GCCGACCGGCTCTGT 264
DB 938 GCTTGGGGGT 997
QY 265 TCCGCGACCGGCTGT 324
DB 998 ACGGTGGCGCGCGGT 1057

QY 325 CCGGATCGGCGCGCGCATTCGACCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 384
DB 1058 GCTTGGGGGT 1117
QY 385 TGTCTGCGGCTGT 444
DB 1118 TGTCTGATCGGCGCGCGGT 1177
QY 445 TCCAGGCGCTCGGCGCGCGCGGATCGACGTCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 504
DB 1178 GGGGCG 1237
QY 505 CCGCGGTGACCGGT 564
DB 1238 GCGGCGCGCATCGGCGGCTTGT 1297
QY 565 GGG 567
DB 1298 AGG 1300

RESULT 11

US-11-254-686-8
; Sequence 8; Application US/11254686
; Publication No. US20060040877A1
; GENERAL INFORMATION:
; APPLICANT: Martin, Christine J.
; APPLICANT: Burns, Lesley S.
; APPLICANT: Graupner, Paul R.
; APPLICANT: Lewer, Paul
; APPLICANT: Voudsen, William A.
; APPLICANT: Wilkinson, Barrie
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Novel Spinosyn-Producing Polyketide Synthases
; FILE REFERENCE: 62069
; CURRENT APPLICATION NUMBER: US/11/254,686
; PRIOR FILING DATE: 2005-10-20
; PRIOR APPLICATION NUMBER: US 60/358075
; PRIOR FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Saccharopolyspora erythraea
US-11-254-686-8

Query Match 9.9%; Score 59.4; DB 11; Length 1685;
Best Local Similarity 47.0%; Pred. No. 0.0003;
Matches 254; Conservative 0; Mismatches 281; Indels 6; Gaps 2;

QY 63 GAGAGCCCGGTTCGACCTGCGGCTTACCGGTTCGACGAGCAAGAACTTGTGCTCCCGTTC 122
DB 1000 GAGAGT 1059
QY 123 CCGGTGCG 182
DB 1060 CCGGTGCGCTTTCGACGAGCGGATCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
QY 183 GTGACCTCGGCGCGGCTCTCCCTGT 242
DB 1120 GAGCGAGCG 1179
QY 243 CCGCGCGCTGT 302
DB 1180 TCTTGTGCG 1239
QY 303 TTTGTGACCGGCGAGT 359
DB 1240 CTGCG 1299
QY 360 ACCGCGCTGT 419

Db 1300 GACGTGAGGCGGACCCGCTCTCTGCGGAGTTGCGCGGCGGACGAGAGACGAG 1359
QY 420 CCCAGCTGAGACCCCTCGGCTGTGTCAAGGCGCTCGGCGGACCGGAGATTCAGTCCC 419
Db 1360 CCGGCGGAGTCCGCGCTGCACTG---AACGCGCACCGGACCTGTGCTCGGAGCGGCTG 1416
QY 480 CTGTGCTGTGTGACCAAGGAGCGCGCGCGCTGACCGTTCGAGACGATTCGAGCC 539
Db 1417 CTGCGGCTGTCAACGCGGAGACCGCGCTTGGCGGCGCGGACCGGACGCGGAGCC 1476
QY 540 CAGGCGATGTGCTGCTGCTGCGCGGCTGTGTGAGTCCCGCGCGGAGTGGGT 599
Db 1477 ACCTTCGCGAGCTGAGGCTGAGCTGCTGTGCGCGGACGCTGCGCGGAGGTGAGC 1536
QY 600 G 600
Db 1537 G 1537

RESULT 12

US-11-052-554A-522/C
; Sequence 522, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 522
; LENGTH: 4146
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-522

Query Match 9.8%; Score 58.6; DB 14; Length 4146;
Best Local Similarity 46.6%; Pred. No. 0.00037;
Matches 224; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 120 GTCCCTGAGGCGGCTGCTCCGACACCGGCGAGAGTGGCTTACGAGCTGCTGAGC 179
Db 2046 GGGGCGGCGGCGCGCGCTGACGCGCGCTGTGCGACCGGCGCGCGCTGAGCGCTTC 1987
QY 180 CCGGTGACCTTGGGCGGCTCTCTGCGCGGAGCGGCTGTGTGACCGGAGCGGCA 239
Db 1986 GCGGTGACCCCAAAATTGCGGCTGTGCGGAGCGGCTTCCGCGCGCGGCTGTC 1927
QY 240 GAAACCCCGGCTGAGACGAGTGTGCGGAGCGGCTTGAAGAAGCGGCGGAGACCTTC 299
Db 1926 GCGGCGGCGGCTGCGCGCGGAGCGCGGCTGTGCTGCGGCTGCGGCTGCGGCTT 1867
QY 300 GTGTTGTGACGCGGAGTGTGCGGCGGAGTGGGCGGCGGCACTTCGACGCGCTGACGCG 359
Db 1866 GCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTTCGACGAGC 1807
QY 360 ACCGCGCTGTGACCTGTGTGTCTGTGCTGCGGCTGCGGAGGCGGCTGTGTGAGC 419
Db 1806 TGGGAGCATGTGTGTGTGTGCGGAGTTC---GCGGCTTTCGCGCGGCTGCGGCTT 1750
QY 420 CCCAGCTGAGACCCCTCGGCTGTGTGCAAGGCGCTGCGGCGGAGATTCAGTCCC 479
Db 1749 GCGGCGGCTGTGCGGCTGTGCGCGGCTGTGCGGCTGTGACCGCGGCTGACCGCT 1690
QY 480 CTGTGCTGTGTGACCAAGGAGCGCGCGCGCTGACCGTTCGAGACGATTCGAGCC 539
Db 1689 CCGGCGGCGGCTCCCGG 1630

QY 540 CAGGCGATGTGCTGCTGCTGCGCGGCTGTGTGAGTCCCGCGGAGTCCCGCGGAGTGGGT 599
Db 1629 CCGGCGGAGGCTGTGCGGCTGTGAGGCTGTGTGCGCGCGCTTCAACGCTCGGAGCGCTT 1570
QY 600 G 600
Db 1569 G 1569

RESULT 13

US-11-228-659-45
; Sequence 45, Application US/11228659
; Publication No. US20060064786A1
; GENERAL INFORMATION:
; APPLICANT: Brugiere, Norbert
; TITLE OF INVENTION: Isopentenyl Transferase Sequences and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 1507
; CURRENT APPLICATION NUMBER: US/11/228,659
; PRIOR FILING DATE: 2005-09-16
; PRIOR APPLICATION NUMBER: 60/610,656
; PRIOR FILING DATE: 2004-09-17
; PRIOR APPLICATION NUMBER: 60/637,230
; PRIOR FILING DATE: 2004-12-17
; PRIOR APPLICATION NUMBER: 60/696,405
; PRIOR FILING DATE: 2005-07-01
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: OsIPT2 coding sequence (018830_1)
; NAME/KEY: CDS
; LOCATION: (1)...(984)
US-11-228-659-45

Query Match 9.7%; Score 58.2; DB 11; Length 984;
Best Local Similarity 47.0%; Pred. No. 0.00061;
Matches 214; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 107 CTTCGCTGCTCCGAGTCCCGCTGCGGCTGCGGACACCGGAGAGTGGCTTACCA 166
Db 216 CGTGCGGACCACTCTGCGGCGGCTGAGCGCGCGGACGCGGACTTACCGCGGAGACTT 275
QY 167 GCTGCGCTGACACCCGCTGACCTTGGGCGGCTCTTCCCTGCGGAGCGGCTCTGTGTGT 226
Db 276 CCGCGGAGGCGGCGGCGGCGGCTGCGCGGCTGCGGCGGCGGCGGCGGCGGCTTCCGCT 335
QY 227 GACCGGAGGCGGAGTACCCCGGCTGAGACGAGTGTGCGGAGCGGCTGGAACAAGCG 286
Db 336 CCGCGGCGGCTGGAACATTTAGTGTGAGGCGCTGTGCGGCGGCGGCGGCGGCGGCTTCT 395
QY 287 CCGGCGGACCGTGTGTGTGACCGGAGTGTGCGGCGGCGGCGGAGTGTGCGGCGGCACTTCA 346
Db 396 CCGGCGGCTGACCTGCTTCTGTGACCGACGCTGCGGCGGAGCTGTGCGGCTGAGTGA 455
QY 347 GCGCGTGAACGCGACCGGCTGTGCACTGTGTCTTCTGCTGCGGCTGCGGAGGCGG 406
Db 456 CAGGCGGCGGCGGCTGTGAC---GACATGTGTGCGGCGGCGGCTGTGTGCGGAGCGGCGG 512
QY 407 TGTGTGACGACACCGAGCTTGAACCTTCCGCTGTGTGCTGAGGCGCTTGGGCGGAGCGG 466
Db 513 CCGGCTTGAACCGCGGCGGAGTACACCGCGGCGGCTGTGCGGCGGCTGTGAGTACCGCA 572
QY 467 GATGCAAGTCCCTGTGTGCTGTGTGACCAAGGAGCGGCGGCGGCTGTGAGAGCA 526
Db 573 GATGCAAGCTTCTGTGTGCGGAGCGGAGGCGGCGGCGGCGGCGGCGGAGAGAGACGACGA 632

QY 527 CGTCGATCCGAGCCGAGGCGATGTCGAGGCTCG 561
DB 633 CCTCTCGCCGCGATGCTCGAGGCGCCGCTGCGCG 667

RESULT 14

US-11-228-659-44
; Sequence 44, Application US/11228659
; Publication No. US20060064786A1
; GENERAL INFORMATION:
; APPLICANT: Brungiere, Norbert
; TITLE OF INVENTION: Isopentenyl Transferase Sequences and
; FILE REFERENCE: 1507
; CURRENT APPLICATION NUMBER: US/11/228,659
; CURRENT FILING DATE: 2005-09-16
; PRIOR APPLICATION NUMBER: 60/610,656
; PRIOR FILING DATE: 2004-09-17
; PRIOR APPLICATION NUMBER: 60/637,230
; PRIOR FILING DATE: 2004-12-17
; PRIOR APPLICATION NUMBER: 60/696,405
; PRIOR FILING DATE: 2005-07-01
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 4470
; TYPE: DNA
; ORGANISM: Oryza sativa

FEATURE:
NAME/KEY: CDS
LOCATION: (1484)...(2470)

FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)

OTHER INFORMATION: OsIPT2 full length sequence (018830_1)

FEATURE:
NAME/KEY: misc_feature

LOCATION: 2876..2877, 2878, 2879, 2880, 2881, 2882, 2883, 2884, 2885,
LOCATION: 2886, 2887, 2888, 2889, 2890, 2891, 2892, 2893, 2894, 2895,
LOCATION: 2896, 2897, 2898, 2899, 2900, 2901, 2902, 2903, 2904, 2905,
LOCATION: 2906, 2907, 2908, 2909, 2910, 2911, 2912, 2913, 2914,
OTHER INFORMATION: n = A,T,C or G

FEATURE:
NAME/KEY: misc_feature

LOCATION: 2915..2916, 2917, 2918, 2919, 2920, 2921, 2922, 2923, 2924,
LOCATION: 2925, 2926, 2927, 2928, 2929, 2930, 2931, 2932, 2933, 2934,
LOCATION: 2935, 2936, 2937, 2938, 2939, 2940, 2941, 2942, 2943, 2944,
LOCATION: 2945, 2946, 2947, 2948, 2949, 2950, 2951, 2952, 2953,
OTHER INFORMATION: n = A,T,C or G

FEATURE:
NAME/KEY: misc_feature
LOCATION: 2954..2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963,
LOCATION: 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973,
LOCATION: 2974, 2975,
OTHER INFORMATION: n = A,T,C or G

US-11-228-659-44

Query Match 9.7%; Score 58.2; DB 11; Length 4470;
Best Local Similarity 47.0%; Pred. No. 0.00044;
Matches 214; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 107 CTTCTGCTCCCGGCCCCCTGCGCGCGGTCCCGACCGCGGACGAGTGCCTTACCA 166
DB 1699 CGTGGGACCACTCTCTGCGGCGGTGACGCCCGACCGACTTCAACCGGAGACTT 1758
QY 167 GCTGCGCTGACACCCGTCGACCTGCGGCGGTCTCCCTGACCGGAGCTGTGATG 226
DB 1759 CCGCGCGAGGCGCGCGCGCGCTGCGCGCGGTCCCGCTCAACCGGCGCGCTCTCCGCTGT 1818
QY 227 GACCGAGCGGACGATACCCCGCGCTGACCGGACGTCGTCCGCGACCGGCTTGAAACAGCG 286
DB 1819 CGCGCGCGGTGCAACATCTACGTGACGAGCGCTGTCGCGCGCGCGCGCGGCGCTTCT 1878

QY 287 CGGAGCGACCGTCGTGTTGTGACACCGGACATGCGCGCGCCCGATCGGCGCGACTTCA 346
DB 1879 CGGCGGTATGACATGCTCTTCTCTGTGACACGATGCGCGCGGACCTGCTCGGTGAT 1938
QY 347 CGCGGTGACGAGCAGCGCCCTGTCACCTGTGATCTCTCTGCTGCGCTCGCGAGGCGCG 406
DB 1939 CAGCGCGCGCGCGGTGAC---GACATGTGTGCGCGCGCGGTGTGCGAGCGCGCGC 1995
QY 407 TGTCTGACGACCCCGACGCTGACACCTCTGCTTGTGATCAAGCGCTTCGCGAGCGCG 466
DB 1996 CGGTTGACGCGCGCGCGGACCTACACCCGCGGCGGTGCGCGCGCATCGGAGCTACCGCA 2055
QY 467 GATCGACGTCCTCCCTGTGCTGTGTCGACCAAGGACGCGCGCGCTGACCGTCCGAGACGA 526
DB 2056 GATCGACGCTACTCTCTGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2115
QY 527 CGTCGATCCGAGCCGAGGCGCATGTCGAGTGGGCTCG 561
DB 2116 CCTCTCGCCGCGATGCTCGAGGCGCGCGGTGCGCG 2150

RESULT 15

US-11-075-185-36
; Sequence 36, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:

APPLICANT: REEVES, CHRISTOPHER D

APPLICANT: JULIEN, BRYAN

TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS

FILE REFERENCE: 010099.03

CURRENT APPLICATION NUMBER: US/11/075,185

PRIOR APPLICATION NUMBER: US 60/551,103

PRIOR FILING DATE: 2004-03-08

PRIOR APPLICATION NUMBER: US 60/568,290

PRIOR FILING DATE: 2004-05-04

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.3

SEQ ID NO 36

LENGTH: 5679

TYPE: DNA

ORGANISM: Sorangium cellulosum

US-11-075-185-36

Query Match 9.7%; Score 58.2; DB 14; Length 5679;
Best Local Similarity 46.4%; Pred. No. 0.00042;
Matches 189; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 172 CTTGAGACCCCGTCGACCTGCGGCGGTCTCCCTGCGCGGACGCGGCTCTGTGTCGACCG 231
DB 3776 CCGAGAGACAGTGTGTGTGTCGCGCGGTGACCGGTGCGCTGCGCGGCGGCTGCGCG 3835
QY 232 GAGCGGACGATACCCCGCGCTGACCGGACGTCGTCCGCGACGCGCTGGAACAGCGCGGG 291
DB 3836 CGGCGCGCGGCTGCGCGGAGCTGTCGAGGCGGTGCGGACGCGGCGCGCGCGCGCGC 3895
QY 292 CGACCTGTGTGTGACCGCGGACGTCGCGCGCGCGCGGATGCGGCGCGCACTTCACGCGC 351
DB 3896 GCGTGTGTGTGACCTGACCGCGCGGCGGACGTCGCGCGCGGTGTCGCGCGCGCGCG 3955
QY 352 TCGACGCGACCGCGCTGTCACCTGTGTGTCTCTGCTGCGCTGCGCGGAGCGGATGTCG 411
DB 3956 CCGTGTGCGGCGCGCTGCGCTGTGTGTGAGGAGGTGCTGCGCGCGCGCGCGCGCGCG 4015
QY 412 TCGACGACCCCGACGCTGACACCTTGTGCTGTGTCACGCGCGCTGCGCGCGGATGCG 471
DB 4016 CCGAGCTCTGTGTGTGACCGGCTGCGCGGTGTGCGGACCGCGCGCGGAGCGGCGGTGAG 4075
QY 472 ACGTCTCCCTGTGTGTGTGACGAGGACGCGCGCGGTGACCGTGTGGAACGACGTCG 531
DB 4076 CGTGTGCGCGCGCGGTGTGTGTGTGCGGCGCGACCGCGCGCGGACGATACCGCGAC 4135

Mon Apr 10 09:27:47 2006

us-10-611-442-2_copy_1_600.rnpbn

Page 9

QY 532 ATCCGGCCCAAGCCATGTCGGTGGGCTCGGCGGGGTGTCGGGCGTG 578
 Db 4136 GCGCGGTCCGGGTCTGACGTGGGGCGCGAGCCGCTGACCGGGCG 4182

Search completed: April 7, 2006, 03:50:12
Job time : 228.15 secs

is Page Blank (uspto)

Qy	1	CCGGGCTTCGCGAAGCGTGAACACCCGGGGGGGGTGAAGTGACTGGGGTACCGGAGTGGGTG	60
Db	1	CCGGGCTTCGCGAAGCGTGAACACCCGGGGGGGGTGAAGTGACTGGGGTACCGGAGTGGGTG	60
Qy	61	AGGAGCGCCCGGTGACCTTCGCGGTCTACCCGTTCCAAAGACAGAACTTCTGAGTCCCGG	120
Db	61	AGGAGCGCCCGGTGACCTTCGCGGTCTACCCGTTCCAAAGACAGAACTTCTGAGTCCCGG	120
Qy	121	TCGCCCTTGGGCGGGGTCCCGGACACACCGCGAGACGAGTGGCGTTTACGAGCTCGCTGGCAAC	180
Db	121	TCGCCCTTGGGCGGGGTCCCGGACACACCGCGAGACGAGTGGCGTTTACGAGCTCGCTGGCAAC	180
Qy	181	CCGTCGACCTTCGGGCGGTCTTCCTTCGCGGAGCGGTCCTGGTGGTGAACCGGAGCGGACG	240
Db	181	CCGTCGACCTTCGGGCGGTCTTCCTTCGCGGAGCGGTCCTGGTGGTGAACCGGAGCGGACG	240

Db 181 CCGTGCACCTCGGCGGCTCCCTCGGCGGAGCGGCTCTGTGTGTGACCGGAGCGGAG 240
Qy 241 TACCCCGGCGCTGAGACGAGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 300
Db 241 TACCCCGGCGCTGAGACGAGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 300
Qy 301 TGTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 360
Db 301 TGTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 360
Qy 361 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 420
Db 361 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 420
Qy 421 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 480
Db 421 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 480
Qy 481 TGTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 540
Db 481 TGTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 540
Qy 541 AGGCGATGT 600
Db 541 AGGCGATGT 600

RESULT 2

US-10-672-396-5
; Sequence 5, Application US/10672396
; Publication No. US20040166567A1
; GENERAL INFORMATION:
; APPLICANT: SANTI, DANIEL V.
; APPLICANT: REID, RALPH C.
; APPLICANT: KODUMAL, SARAH J.
; APPLICANT: JAYARAM, SEBASTIAN
; TITLE OF INVENTION: SYNTHETIC GENES
; FILE REFERENCE: 300622010900
; CURRENT APPLICATION NUMBER: US/10/672,396
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US 60/414,085
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 9510
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-10-672-396-5

Query Match 31.5%; Score 189; DB 7; Length 9510;

Best Local Similarity 59.5%; Pred. No. 1.7e-37;
Matches 359; Conservative 0; Mismatches 235; Indels 9; Gaps 2;

Qy 1 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 60
Db 2582 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 2641
Qy 61 AGGCGCGGCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAG 120
Db 2642 ATGCGCGGCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAG 2701
Qy 121 TCC--CCCTGT 177
Db 2702 TTCCGAGAGGT 2761
Qy 178 ACCCGGT 237
Db 2762 GTGAGCGGAGT 2821

Qy 238 CAGTACCCCGGCGGCTGTGTGACGAGAGT 297
Db 2822 GTGTACCATCTGT 2881
Qy 298 TCGT 357
Db 2882 TTTTGACATGT 2941
Qy 358 GCACCGGCGGCTGT 417
Db 2942 CTGATGT 2995
Qy 418 ACCCGAGCGGCGGCTGT 477
Db 2996 ATCCGAGCTGT 3055
Qy 478 CCGT 537
Db 3056 CGCTGT 3115
Qy 538 CCGGCGGCGGCTGT 597
Db 3116 CCGGCGGCGGCTGT 3175
Qy 598 GTG 600
Db 3176 GCG 3178

RESULT 3

US-10-205-032-17
; Sequence 17, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianhu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205,032
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-17

Query Match 20.7%; Score 124; DB 5; Length 4725;

Best Local Similarity 55.0%; Pred. No. 1.9e-21;
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;

Qy 1 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 54
Db 2642 CCGGCTGTGTGACCGGCGGAGTGTCCGCGAGCGGCTGTGAAACAGCGGCGGCGAGCGGCTCG 2701
Qy 55 TGGGT 114
Db 2702 GGGCGGCGGCTGT 2761
Qy 115 TCCCGGCTGT 171
Db 2762 TGGAGCGGCGGCTGT 2821
Qy 172 CTTGAGACCGGCTGT 231
Db 2822 GCTGAGCGGCTGT 2881
Qy 232 GAGGCGGAGT 291
Db 2882 GCCGAGGAGT 2941

QY 292 CGACCGTGTGTGTGACACCGGCACTG-----CGGCGCGGATCGGCGCG 339
DB 2942 CCGAGGTGCGCTGTGTGACCGGTGAGCGCAACGAGACCGGCGCGGCGCGCTGT 3001
QY 340 CACTGACCGCGGTGTGACGAGACCGGCTGTCTACTGTGTCTGTCTGTCTGT 390
DB 3002 TGTCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3061
QY 391 -----CGTCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
DB 3062 GTGCGCGGACCGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3121
QY 445 TCGAGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 504
DB 3122 CCGAGGCGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3181
QY 505 CCGCGGTGACCGGTGTGACGAGACGTCGATCCGCGCGGCGGCGGCGGCGGCGG 564
DB 3182 TCTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3241
QY 565 GGGTGTGAGCGGTGTGAGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 3242 GGGTGTGAGCGGTGTGAGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 3277

RESULT 4

US-10-205-032-1
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Seafra, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: microsporidia carbonacea subspecies aurantiaca
US-10-205-032-1

Query Match 20.7%; Score 124; DB 5; Length 60196;
Best Local Similarity 55.0%; Pred. No. 1.1e-21;
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;

QY 1 CCGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 54
DB 38933 CCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 38992
QY 55 TGGGTGAGGAGCGCGCGGTGACCTGCGCGGTCTACCGGTCTCAACGACGAACTTGTGGC 114
DB 38993 GGGCGCGCGCGCGGTGACCTGCGCGGTCTACCGGTCTCAACGAGCGGAACTTGTGGC 39052
QY 115 TCCCGGTCCCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 171
DB 39053 TGGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39112
QY 172 CCGTGGACCGCGCGGTGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 221
DB 39113 GCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39172
QY 232 GAGCGGCACTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 291
DB 39173 GCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39232
QY 292 CGACCGGTGTGTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 339
DB 39233 CCGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39292

QY 340 CACTGACCGCGGTGTGACGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
DB 39293 TGTCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39352
QY 391 -----CGTCCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
DB 39353 GTGCGCGGACCGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39412
QY 445 TCGAGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 504
DB 39413 CCGAGGCGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39472
QY 505 CCGCGGTGACCGGTGTGACGAGACGTCGATCCGCGCGGCGGCGGCGGCGGCGG 564
DB 39473 TCTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 39532
QY 565 GGGTGTGAGCGGTGTGAGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 39533 GGGTGTGAGCGGTGTGAGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 39568

RESULT 5

US-10-203-295-35
; Sequence 35, Application US/10203295
; Publication No. US20040115762A1
; GENERAL INFORMATION:
; APPLICANT: Zolchev, Sergey Borisovich
; APPLICANT: Sekurova, Olga Nikolayivna
; APPLICANT: Pjaeravik, Bpsen
; APPLICANT: Brautaset, Tryve
; APPLICANT: Strom, Arne Reidar
; APPLICANT: Valla, Svein
; APPLICANT: Ellingsen, Trond Erling
; APPLICANT: Sletta, Hvard
; APPLICANT: Gulliksen, Ole-Martin
; TITLE OF INVENTION: Novel genes encoding a nystatin polyketide synthase and their
; FILE REFERENCE: 1181-265
; CURRENT APPLICATION NUMBER: US/10/203,295
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/GB 01/00509
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: GB 0002840.7
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: GB 0008786.6
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: GB 0009387.2
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 125401
; TYPE: DNA
; ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-35

Query Match 17.2%; Score 103.2; DB 7; Length 125401;
Best Local Similarity 53.5%; Pred. No. 1.3e-16;
Matches 291; Conservative 0; Mismatches 238; Indels 15; Gaps 3;

QY 72 GTGACCTGCGCGGTGTGACCGGTCTCAACGAGACGAACTTGTGCTCCCGGTCGCGGCG 131
DB 43840 GCGACCACTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 43899
QY 132 CCGGTCCCGGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 191
DB 43900 ACATGCGGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 43959
QY 192 GGGCGGTCTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 248
DB 43960 AGCGCGGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 44019
QY 249 GCGTGGACGGAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 308

Db 44020 GTGCGCAGATGTCGCGCCCTGCGCCGAGACCGCCGCGGTGAGATGTCGCCGCGAC 44079
Qy 309 ACCGCGGAGTCGCGCGCCCGGATGCGGCGCGCACTGACACCGCTGACGCGACCGCTG 368
Db 44080 GGCACCGGACCGCGCGCGCTGCGCCCTGCTGCTACCGAAGGCGCGACGACCGCCCG 44139
Qy 369 TCCATGTCGTCCTCTGTCGCGCTGCGCGCGCGCGCGCTGTCGACGA-----C 419
Db 44140 ACCGCGGTGTCCTCTGTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44199
Qy 420 CCCAGCTGACACCTGCGGTGTCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
Db 44200 GGCACACCG 44259
Qy 480 CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 536
Db 44260 CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 44319
Qy 537 GCGCAGGCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 596
Db 44320 GCGCAGGCG 44379
Qy 597 GGTG 600
Db 44380 GCGG 44383

RESULT 6

US-10-205-032-15
; Sequence 15, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianhu
; APPLICANT: Stafla, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 11238
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-15

Query Match 16.7%; Score 100.2; DB 5; Length 11238;
Best Local Similarity 52.9%; Pred. No. 1.2e-15;
Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1;

Qy 143 CACCGGCGACGATGCGCTTACCAAGCTGCGCTGCGACCCCGTGCACCTCGGCGGTCTC 202
Db 2949 CACCTGAAACCGTTGCCGACCGACGACGCCCGCTTCAACACCGCTGCTGCTGAT 3008
Qy 203 CCGGCGGAGCGGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 262
Db 3009 GGTGCGGAGCGGCTGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3068
Qy 263 GGTGCGGAGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 322
Db 3069 GGTGCGGAGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3128
Qy 323 GCGCGGATGCG 382
Db 3129 GCTCTGACCTGCG 3188
Qy 383 TCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 428
Db 3189 CTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3248
Qy 429 -GACACCGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 487

Db 3249 CGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3308
Qy 488 GGTGACGAGGAGCG 547
Db 3309 GGTGACGAGGAGCG 3368
Qy 548 GGTGAGGAGCG 600
Db 3369 GGTGAGGAGCG 3421

RESULT 7

US-10-819-386A-1/c
; Sequence 1, Application US/10819386A
; Publication No. US2005089884A1
; GENERAL INFORMATION:
; APPLICANT: Korea Advanced Institute of Science and Technology
; APPLICANT: Shanghai Jiaotong University
; APPLICANT: Lee, Sang Yup
; APPLICANT: DENG, ZIXIN
; APPLICANT: CHEN, SHI
; APPLICANT: JIONG, Ki Jun
; APPLICANT: ZHOU, XIUPEN
; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polypeptides
; FILE REFERENCE: P003-B015
; CURRENT APPLICATION NUMBER: US/10/819,386A
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: KR10-2003-0074035
; PRIOR FILING DATE: 2003-10-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 138203
; TYPE: DNA
; ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1

Query Match 16.7%; Score 100.2; DB 9; Length 138203;
Best Local Similarity 53.7%; Pred. No. 6.9e-16;
Matches 341; Conservative 0; Mismatches 258; Indels 36; Gaps 5;

Qy 2 CCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 61
Db 92504 CCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92445
Qy 62 GGAAGCGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 121
Db 92444 GGAAGCGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92385
Qy 122 CCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 179
Db 92384 GCG 92325
Qy 180 CCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 232
Db 92324 GCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92265
Qy 233 AGCGGAGTCACCG 292
Db 92264 GCGTCACCG 92205
Qy 293 GACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 349
Db 92204 CCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92145
Qy 350 CCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 400
Db 92144 CCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92085
Qy 401 -----GCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 448
Db 92084 GCGAGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 92025

QY 449 GGCGCTGGGGCAGGCCGGGATTCACACGGTCCCCCTGTGGCTGTGTATCCAGGGAGACGGCCGGC 508

Db 92024 GGCCCTCTGTAGACGCCGGGGGTTCACACGGCCCCCTGTGTGTCTGTATACCCCGGGAGACCGCTCC 91965

QY 509 CGTACACCGT---CGAGACGACGTCATTCGGGCCAGAGCCATGTGCGTGGGCTGGAGCC 565

Db 91364 GGTGGCGGGATTCCAGGCGCTCAACGGCCCCCGCCAGGCCCGCGTGTGGGGCCTGTGGCCG 91905

QY 566 GGTGTGGGCGTGTGAGTCCGCCGCCCGATGGGGTG 600

Db 91804 GGTGGCGGCCCTCGAACAACCCCGCCGTGTGGGGCG 91870

RESULT 8

```

US-10-042665A-3
? Sequence 3: Application US/10042665A
? Publication No. US20050053927A1
? GENERAL INFORMATION:
? APPLICANT: Schnupp, Thomas
? APPLICANT: Toupet, Christine
? APPLICANT: Engel, Nathalie
? TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
? FILE REFERENCE: 4-21001/E/C1
? CURRENT APPLICATION NUMBER: US/10/042,665A
? CURRENT FILING DATE: 2002-01-09
? PRIOR APPLICATION NUMBER: 09/102,744
? PRIOR FILING DATE: 1999-03-24
? PRIOR APPLICATION NUMBER: PCT/EP97/04495
? PRIOR FILING DATE: 1997-08-18
? NUMBER OF SEQ ID NOS: 9
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 3
? LENGTH: 53799
? TYPE: DNA
? ORGANISM: Amycolatopsis mediterranei
US-10-042-665A-3

```

[illegible]

QY	599	TG	600
Db	14079	CG	14080

RESULT 9

```

US-09-861-289-30
Sequence 30, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIORITY APPLICATION NUMBER: 09/105,537
PRIORITY FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

```

Query Match	14.9%	Score 89.4	DB 3	Length 13842
Best Local Similarity	52.6%	Pred. No. 5e-13		
Matches 317	Conservative 0	Mismatches 211	Indels 15	Gaps 5
QY 1	CCGCGCTCCGACAGGCGTACACCCGGGGGGTGAAGTGCAGCTGGCGTACCGCAGATG---G	57		
Db 5849	CTCTGCTCGCGAGGCGTGGGTCAACGGGCTTCCCGTGATGACATTCCGCTCTGGCCG	5908		
QY 58	GTMAGGAGACGCCCGCGTTCGACTTCGCGGTCTAACCCGTTCCAAACGACAGAACTTCTGGCTTC	117		
Db 5909	CCAGGCGCTCCGCGCCCGCGTGTGACCACTAACGCTTCCAGGCCGACCGCGCTACCTGGCTCG	5968		
QY 118	CGGTCCCTCGGGGCGGGGTCCCCGACACGCGGCGACGAGTGGAGCTTTACAGCTCGCTGGC	177		
Db 5969	AGAACTACTCCGCG---GCCCTGGCCACCGGCGACATGAGCTTACCGCATTCATCACTGA	6028		
QY 178	ACCCGCTGACCTCGGGCGGTCTCTCTGGCCGAGCGAGTCTTGNTGTGACCGGAGCGG	237		
Db 6026	AGGCGCTCCGCGCGCGCGAGAGGGTCCGAGGCGACCG---CTGTCTCGGCGCGTGGCTCG	6082		
QY 238	CAATTACCCCGCGCTGACCGAAGTGTCCGCGACCGGCTTGGAAACAGCGCGGGCGACCG	297		
Db 6083	CGCTCACGCGGAGGACCACTTCGCGAGGCGCGCGCGGTGTCACGCGCGTGTGCAGC	6142		
QY 298	TCGTGTGTGTCACCGGCGAGTCCGCGCGCGGATTCGGCGCGGCACTGACGCGGTGCAGC	357		
Db 6143	CCGGGCGAAGGTCTGAGGTGTACCGCGCGGGCGGACGACGACCGTGAAGCCCTTCGCG	6202		
QY 358	GACCGCGCTGTTCACCTGTGTCTCTGTGCTCGCGCTCGCGAGGCGGTGTGTGACG	417		
Db 6203	---CCGGGCTCACCGACGTGACGACCGGTGACGGCTTCAACCGCGTGTCTCGCTCTCG	6259		
QY 418	ACCCGAGCGCTGACACCCCTCGCGTGTGTCCAGGCGCTCGGCGGACCGCGGATGAGCGTCC	477		
Db 6260	ACGAGACTCTGTAACGCGACAGGTGACTTGGGTTCAGGCGGCTTCGGGACGCGCGGAATCAAGCGC	6319		
QY 478	CCCTGTGCTGTGTGACACAGGAGCGCGCGCGCTGTGACG---TCGAGACGACGTTCGATC	534		
Db 6320	CCCTGTGTCTCGTCAACCGAGGCGGCGGTCTCGTCCGAGCGTTCGACACCCCGCGGACG	6379		
QY 535	CGGCGCCAGGCGATGATGTGCGTGGGCTTCGCGCGGCTGTGGCGTGTGAGTCCCCCGCCGCT	594		
Db 6380	CCGACCGGGGCGAAGCTCTGTGGGGCTTCGGCGCGCTGTGCGCTTGTGACACCCCGAAGCT	6439		

QY 595 GGG 597
Db 6440 GGG 6442

RESULT 10
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match 14.9%; Score 89.4; DB 3; Length 13842;
Best Local Similarity 52.6%; Pred. No. 5e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCCGCGAGCGGTACACCCGCGGGGTGAGGTGCTGACCTGCGCTACCGCAGTGG---G 57
Db 5849 CCTGCTCCGCGAGCGGTGAGTCAACGGGCTTCCCGTGCATGACCTTCCCTGCGCCG 5908
QY 58 GTGAGGAGACGCGCGGTGACCTGCGGTCTACCCGTTCAACGACAGAACTTTGCGCTCC 117
Db 5909 CCACGGCTCCCGCGCGGTGCTGCGCTTCCAGGCGGAGGCTACTGCGCTCG 5968
QY 118 CGGTCCCTGCGCGGGTCCCGGACACCGGCGACGAGTGGCGTTTACCACTCGCTGCGC 177
Db 5969 AGAACACTCCCGC---GCCCTGCGCACCGGCGACGACTGCGCGCTACCGCATCGACTGGA 6025
QY 178 ACCCGTGCACCTGCGGGCGGTCTCCCTGCGCGGACGAGCGGTCTGTTGTTGACCCGAGCG 237
Db 6026 AGCGCTCCCGCGCGCGAGGGGTCCAGCGCACCG---CTGTCCCGCGCTGCGCTCG 6082
QY 238 CAGTACCCCGCGCTGACGAGACGTGTCGCGACGAGCGCTGGAACAGCGCGGGCGACCG 297
Db 6083 CGGTACCGCGGAGGACCACTCCGCGCAGGCGCGCGCGCTGCTACCGCGCTGTCGACG 6142
QY 298 TCGTGTGTGACACCGCGACGTGCGCGCGCGCGATCGCGCTGACCTGACCGCGCTGACG 357
Db 6143 CCGGCGGAGAGTGCAGAGTGTGACGCGCGGGCGGACGACGACCTGTAGAGCGCTCGCG 6202
QY 358 GCAACGCGCTGTCCACTGTGTCTCTGCTGCGCGCTGCGAGGCGGCGGTGTGTCGACG 417
Db 6203 ---CCCGCTACCGCACTGACGACCGGTGACCGGCTTACCGGCGGTGTGCTGCTCG 6259
QY 418 ACCCAGCTGACACCTCGCGTGTGTCAGGCGCTGCGCGACCGCGGATTCAGCGTCC 477
Db 6260 ACGACTGTACCGGAGGTGCTGCGGTTCAGGCGCTGCGGACCGCGGATTCAGGCGCG 6319
QY 478 CCTGTGCTGTGACACGAGGACGCGCGCGCGCTGACCG---TCGAGACGACGTGATC 534
Db 6320 CCTGTGCTGTGACACGAGGCGCGGCTGCTGCGGACGACGTCTGACACCGCGCGACG 6379
QY 535 CGGCGGAGGACATGTGTGCGTGTGCGCGCGGCGGTGCGCGTGTGAGTGTGCGCGCGCGGT 594
Db 6380 CGGACCGGCGCATGTCTGTGGGCGCTCGCGCGGTGTGCGCTTGTGACACCGCGGACGCT 6439
QY 595 GGG 597

Db 6440 GGG 6442

RESULT 11
US-09-988-384B-30
; Sequence 30, Application US/0988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

Query Match 14.9%; Score 89.4; DB 3; Length 13842;
Best Local Similarity 52.6%; Pred. No. 5e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCCGCGAGCGGTACACCCGCGGGGTGAGGTGCTGACCTGCGCTACCGCAGTGG---G 57
Db 5849 CCTGCTCCGCGAGCGGTGAGTCAACGGGCTTCCCGTGCATGACCTTCCCTGCGCCG 5908
QY 58 GTGAGGAGACGCGCGGTGACCTGCGGTCTACCCGTTCAACGACAGAACTTTGCGCTCC 117
Db 5909 CCACGGCTCCCGCGCGGTGCTGCGCTTCCAGGCGGAGGCTACTGCGCTCG 5968
QY 118 CGGTCCCTGCGCGGGTCCCGGACACCGGCGACGAGTGGCGTTTACCACTCGCTGCGC 177
Db 5969 AGAACACTCCCGC---GCCCTGCGCACCGGCGACGACTGCGCGCTACCGCATCGACTGGA 6025
QY 178 ACCCGTGCACCTGCGGGCGGTCTCCCTGCGCGGACGAGCGGTCTGTTGTTGACCCGAGCG 237
Db 6026 AGCGCTCCCGCGCGCGAGGGGTCCAGCGCACCG---CTGTCCCGCGCTGCGCTCG 6082
QY 238 CAGTACCCCGCGCTGACGAGACGTGTCGCGACGAGCGCTGGAACAGCGCGGGCGACCG 297
Db 6083 CGGTACCGCGGAGGACCACTCCGCGCAGGCGCGCGCGCTGCTACCGCGCTGTCGACG 6142
QY 298 TCGTGTGTGACACCGCGACGTGCGCGCGCGCGATCGCGCTGACCTGACCGCGCTGACG 357
Db 6143 CCGGCGGAGAGTGCAGAGTGTGACGCGCGGGCGGACGACGACCTGTAGAGCGCTCGCG 6202
QY 358 GCAACGCGCTGTCCACTGTGTCTCTGCTGCGCGCTGCGAGGCGGCGGTGTGTCGACG 417
Db 6203 ---CCCGCTACCGCACTGACGACCGGTGACCGGCTTACCGGCGGTGTGCTGCTCG 6259
QY 418 ACCCAGCTGACACCTCGCGTGTGTCAGGCGCTGCGCGACCGCGGATTCAGCGTCC 477
Db 6260 ACGACTGTACCGGAGGTGCTGCGGTTCAGGCGCTGCGGACCGCGGATTCAGGCGCG 6319
QY 478 CCTGTGCTGTGACACGAGGACGCGCGCGCGCTGACCG---TCGAGACGACGTGATC 534
Db 6320 CCTGTGCTGTGACACGAGGCGCGGCTGCTGCGGACGACGTCTGACACCGCGCGACG 6379
QY 535 CGGCGGAGGACATGTGTGCGTGTGCGCGCGGCGGTGCGCGTGTGAGTGTGCGCGCGCGGT 594
Db 6380 CGGACCGGCGCATGTCTGTGGGCGCTCGCGCGGTGTGCGCTTGTGACACCGCGGACGCT 6439
QY 595 GGG 597

Db 6440 GGG 6442

```
RESULT 12
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIORITY FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30
```

Query Match 14.9%; Score 89.4; DB 3; Length 13842;
Best Local Similarity 52.6%; Pred. No. 5e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

```
QY 1 CGCGCTCCGCGGAGGTACACCGGGGGGTGAGGTGACCTGCGCTACCGCAGTG---G 57
Db 5849 CTTCCGCTCCGCGGAGGTGAGGTGACCGGGCTTCCGCGGAGTACCTGCTCTCCGCG 5908
QY 58 GTGAGGAGCGCCCGGTGACCTGCGCTTACCCGTTCAACGACAGAACTTGTGCTCC 117
Db 5909 CCAAGGCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 5968
QY 118 CGGTCCCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 177
Db 5969 AGAACAATCCCGGCG---GCCCTGGCCACCGGCGACGACCTGCGCTACCGCATGACCTGA 6025
QY 178 ACCCGGTGACCTGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 237
Db 6026 AGCGCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6082
QY 238 CAGTACCCCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 297
Db 6083 CCGTCAAGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6142
QY 298 TCGTGTGTGACCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 357
Db 6143 CCGGGGCGAAGGTGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6202
QY 358 GCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417
Db 6203 ---CCGCGCTCAACGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6259
QY 418 ACCCGAGCTGAGCACTGCGGTTGATCCAGGCGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCG 477
Db 6260 ACGAGACTGTAACGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6319
QY 478 CCTGTGCTGTGACCAAGGAGCGCGCGCGCTGACCG---TCGAGAGCAAGTGTGATC 534
Db 6320 CCTGTGCTGTGACCAAGGAGCGCGCGCGCTGACCG---TCGAGAGCAAGTGTGATC 6379
QY 535 CGGCGGAGCGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594
Db 6380 CCGAGCGGAGCGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6439
QY 595 GGG 597
```

Db 6440 GGG 6442

```
RESULT 13
US-10-271-889-30
; Sequence 30, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; PRIORITY FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-30
```

Query Match 14.9%; Score 89.4; DB 6; Length 13842;
Best Local Similarity 52.6%; Pred. No. 5e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

```
QY 1 CGCGCTCCGCGGAGGTACACCGGGGGGTGAGGTGACCTGCGCTACCGCAGTG---G 57
Db 5849 CTTCCGCTCCGCGGAGGTGAGGTGACCGGGCTTCCGCGGAGTACCTGCTCTCCGCG 5908
QY 58 GTGAGGAGCGCCCGGTGACCTGCGCTTACCCGTTCAACGACAGAACTTGTGCTCC 117
Db 5909 CCAAGGCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 5968
QY 118 CGGTCCCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 177
Db 5969 AGAACAATCCCGGCG---GCCCTGGCCACCGGCGACGACCTGCGCTACCGCATGACCTGA 6025
QY 178 ACCCGGTGACCTGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 237
Db 6026 AGCGCTCCGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6082
QY 238 CAGTACCCCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 297
Db 6083 CCGTCAAGCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6142
QY 298 TCGTGTGTGACCGGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 357
Db 6143 CCGGGGCGAAGGTGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6202
QY 358 GCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417
Db 6203 ---CCGCGCTCAACGAGGTGACCTGCGCTTACCCGTTCAACGAGCGCTTACCTGCTCC 6259
QY 418 ACCCGAGCTGAGCACTGCGGTTGATCCAGGCGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCG 477
Db 6260 ACGAGACTGTAACGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6319
QY 478 CCTGTGCTGTGACCAAGGAGCGCGCGCGCTGACCG---TCGAGAGCAAGTGTGATC 534
Db 6320 CCTGTGCTGTGACCAAGGAGCGCGCGCGCTGACCG---TCGAGAGCAAGTGTGATC 6379
QY 535 CGGCGGAGCGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594
```

Db 6380 CCGACCGGCGCATGCTCTGAGGGGCTCGGCGCGCTGCTGCGCTTGAGACACCCGAAAGCT 6439
QY 595 GGG 597
Db 6440 GGG 6442

RESULT 14

US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 14.9%; Score 89.4; DB 3; Length 36778;
Best Local Similarity 52.6%; Pred. No. 4,1e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCGCCGAGCGTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAATG---G 57
Db 7590 CCGGCTCGCCGAGCGTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAATG---G 7649
QY 58 GTAGAGGACGCGCGGTGACCTGCGCGGTCTACCGCTTCAACGACAGAACTTTGAGCTCC 117
Db 7650 CCACGAGCTCCGCGCGCGGTGACCTGCGCGGTCTACCGCTTCAACGAGCGCTTACCTGCG 7709
QY 118 CGGTCCCGCTGAGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 177
Db 7710 AGAACAATCCCGCC---GCCCTGCGCAACCGCGAGACTGCGCTTACCGCAATGAG 7766
QY 178 ACCCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 237
Db 7767 AGCGCTCCCGCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 7823
QY 238 CAGTACCCCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 297
Db 7824 CCGTACCGCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 7883
QY 298 TCGTGTGTGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 357
Db 7884 CCGGAGGAGAGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCG 7943
QY 358 GCAACCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCG 417
Db 7944 ---CCCGCTCACCGCACTGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 8000
QY 418 ACCCAAGCTGACACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCG 477
Db 8001 ACAGACTGTACCGAGGTGCGCTGAGTCAAGCGCTGCGCGGTGACCTGCGCGGTGAC 8060
QY 478 CCTGTGTGTGACCAAGGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 534
Db 8061 CCTGTGTGTGACCAAGGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 8120
QY 535 CGGCGGAGCATGCTGAGGTGCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 594
Db 535 CGGCGGAGCATGCTGAGGTGCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 594

Db 8121 CCGACCGGCGCATGCTCTGAGGGGCTCGGCGCGCTGCTGCGCTTGAGACACCCGAAAGCT 8180
QY 595 GGG 597
Db 8181 GGG 8183

RESULT 15

US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 14.9%; Score 89.4; DB 3; Length 36778;
Best Local Similarity 52.6%; Pred. No. 4,1e-13;
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCGCCGAGCGTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAATG---G 57
Db 7590 CCGGCTCGCCGAGCGTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAATG---G 7649
QY 58 GTAGAGGACGCGCGGTGACCTGCGCGGTCTACCGCTTCAACGACAGAACTTTGAGCTCC 117
Db 7650 CCACGAGCTCCGCGCGCGGTGACCTGCGCGGTCTACCGCTTCAACGAGCGCTTACCTGCG 7709
QY 118 CGGTCCCGCTGAGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 177
Db 7710 AGAACAATCCCGCC---GCCCTGCGCAACCGCGAGACTGCGCTTACCGCAATGAG 7766
QY 178 ACCCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 237
Db 7767 AGCGCTCCCGCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 7823
QY 238 CAGTACCCCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 297
Db 7824 CCGTACCGCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 7883
QY 298 TCGTGTGTGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 357
Db 7884 CCGGAGGAGAGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCG 7943
QY 358 GCAACCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCG 417
Db 7944 ---CCCGCTCACCGCACTGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGAC 8000
QY 418 ACCCAAGCTGACACCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTGCGCG 477
Db 8001 ACAGACTGTACCGAGGTGCGCTGAGTCAAGCGCTGCGCGGTGACCTGCGCGGTGAC 8060
QY 478 CCTGTGTGTGACCAAGGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 534
Db 8061 CCTGTGTGTGACCAAGGACCGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 8120
QY 535 CGGCGGAGCATGCTGAGGTGCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 594
Db 8121 CCGACCGGCGCATGCTGAGGTGCTGCGCGGTGACCTGCGCGGTGACCTGCGCGGTGACCTG 8180

Qy	595	GGG	597
Db	8181	GGG	8183

Search completed: April 7, 2006, 04:35:24
Job time : 424.485 secs

inis Page Blank (uspto)

CC	15-Apr-2004	(first entry)
DE	M. megalomicea	cosmid KOS205-57-2.3B
XX	SEQ ID NO:2.	
XX	ds, polyketide; enzyme; MegR; MegF; MegK; MegCIV; MegCV; MegBVI; MegBIIT.	
XX	MegJ; MegM; megosamine; megalomicin.	
XX	Micromonospora megalomicea.	
XX	WO2004003169-A2.	
XX	08-JAN-2004.	
XX	30-JUN-2003; 2003WO-US020681.	
XX	28-JUN-2002; 2002US-0393016P.	
XX	(KOSA-) KOSAN BIOSCIENCES INC.	
XX	Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H;	
XX	WPI; 2004-203379/19.	
XX	Novel isolated, purified, or recombinant nucleic acid comprising	
XX	polyketide modifying gene, there gene encodes polyketide modifying enzyme	
XX	e.g., MegR, MegK, or MegM enzymes useful for producing modified	
XX	polyketide.	
XX	Example 2; SEQ ID NO 2; 51pp; English.	
XX	The invention relates to a novel isolated, purified, or recombinant	
XX	nucleic acid (I) comprising a polyketide modifying gene, where the gene	
XX	encodes a polyketide modifying enzyme chosen from MegR, MegF, MegK,	
XX	MegCIV, MegCV, MegBVI, MegBIIT, MegJ, and MegM enzymes. A method of the	
XX	invention is useful for producing a modified polyketide, which involves	
XX	culturing a recombinant cell comprising the recombinant nucleic acid	

under conditions in which the cell expresses a product of a gene encoded by the nucleic acid under conditions in which the unmodified polyketide is present, and producing the modified polyketide. The cell produces megalomycin, and can attach megalamine to a polyketide, where the cell, it is naturally occurring non-recombinant state cannot produce megalamine. The present sequence contains downstream megalomycin modification enzyme genes.

Sequence 17596 BP; 2325 A; 5966 C; 6604 G; 2701 T; 0 U; 0 Other;

Query Match 100.0%; Score 600; DB 12; Length 17596;

Best Local Similarity 100.0%; Pred. No. 4.3e-85;

Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCGGCGCTCCGCGAGCGTACACCCCGGGGGTGGAGGTGCACTGGCCGTACCCGCACTGGGTG 60
D 1 CCGGCGCTCCGCGAGCGTACACCCCGGGGGTGGAGGTGCACTGGCCGTACCCGCACTGGGTG 60
QY 61 AGGAGCGCCCGGTGCACTGCGGGGTACCCGTTCCAAACGACAGAACTTCTGGCTCCGG 120
D 61 AGGAGCGCCCGGTGCACTGCGGGGTACCCGTTCCAAACGACAGAACTTCTGGCTCCGG 120
QY 121 TCCGCTGGGCGGGGTCCCGCAACCGGCGACAGTGGCGTTACCAAGCTCGCTGGCAAC 180
D 121 TCCGCTGGGCGGGGTCCCGCAACCGGCGACAGTGGCGTTACCAAGCTCGCTGGCAAC 180
QY 181 CCGTGCACCTGGGCGGTCTCTCTGGCCCGAACGGGTCTCTGGTGTACCGGACCGGACG 240
D 181 CCGTGCACCTGGGCGGTCTCTCTGGCCCGAACGGGTCTCTGGTGTACCGGACCGGACG 240
QY 241 TACCCCGGCGCTGACGAGCGTGTGCGCGACCGGCGCTGGAACAGCGCGGGCGACCGTGC 300
D 241 TACCCCGGCGCTGACGAGCGTGTGCGCGACCGGCGCTGGAACAGCGCGGGCGACCGTGC 300
QY 301 TGTGTGACACCGCGAGTGCAGCGCCCGCAATCGAGCGCGCACTGCACCGCTGCACGCGCA 360
D 301 TGTGTGACACCGCGAGTGCAGCGCCCGCAATCGAGCGCGCACTGCACCGCTGCACGCGCA 360
QY 361 CCGGCTGTGCACTGTGTCTCTGTGCTGGCGGTTCGCCGAGGGCGGTCTGTCCAGCAC 420
D 361 CCGGCTGTGCACTGTGTCTCTGTGCTGGCGGTTCGCCGAGGGCGGTCTGTCCAGCAC 420
QY 421 CCAAGCTGACACCTCGCGTGTGTCAGAGCGCTCGAGCGACCGGAGTCGACGTCGCC 480
D 421 CCAAGCTGACACCTCGCGTGTGTCAGAGCGCTCGAGCGACCGGAGTCGACGTCGCC 480
QY 481 TGTGCTGTGTACCAAGGACCGCGCGCGCTGTGACCGTGCAGACGATCCGAGCC 540
D 481 TGTGCTGTGTACCAAGGACCGCGCGCGCTGTGACCGTGCAGACGATCCGAGCC 540
QY 541 AGGCGATGTGTGTGGGTTCGAGCGGGGTGTGGGTGTGGAATCCCCCGCGTGGGGTG 600
D 541 AGGCGATGTGTGTGGGTTCGAGCGGGGTGTGGGTGTGGAATCCCCCGCGTGGGGTG 600
```

RESULT 2

AAE30757 ID AAE30757 standard; DNA; 47981 BP.

AAE30757;

21-JUN-2001 (first entry)

Micromonospora megalomicea megalomycin biosynthetic gene cluster.

Megalomycin; meg gene; polyketide synthase; antibiotic; motilide;

antiparasitic; de.

Micromonospora megalomicea subsp. nigra.

Key Location/Qualifiers
CDS Complement(1..144)
/*tag= a

```
FT /partial  
FT /gene= "megBVI(megT)" /product= "TDP-4-keto-6-  
FT /note= "encodes AAB82201"  
FT 928..2061  
FT /*tag= b  
FT /gene= "megDVI"  
FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
FT /note= "encodes AAB82202"  
FT 2072..3382  
FT /*tag= c  
FT /gene= "megDI"  
FT /product= "TDP-megosamine glycosyltransferase"  
FT /note= "eryCI, DmrV homologue, encodes AAB82203"  
FT 3462..4634  
FT /*tag= d  
FT /gene= "megI"  
FT /product= "mycarose O-acyltransferase"  
FT /note= "encodes AAB82204"  
FT 4651..5775  
FT /*tag= e  
FT /gene= "megDII"  
FT /product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"  
FT /note= "eryCI, DmrV homologue, encodes AAB82205"  
FT 5822..6595  
FT /*tag= f  
FT /gene= "megDIII"  
FT /product= "daunosaminyl-N-dimethyltransferase"  
FT /note= "eryCVI homologue, encodes AAB82206"  
FT 6592..7197  
FT /*tag= g  
FT /gene= "megDIV"  
FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"  
FT /note= "eryBVIII, dnmV homologue, encodes AAB82207"  
FT 7220..8206  
FT /*tag= h  
FT /gene= "megDV"  
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
FT /note= "eryVII, dnmV homologue, encodes AAB82208"  
FT complement(8228..9220)  
FT /*tag= i  
FT /gene= "megBII-1(megDVII)"  
FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"  
FT /note= "encodes AAB82209"  
FT complement(9226..10479)  
FT /*tag= j  
FT /gene= "megBV"  
FT /product= "TDP-mycarose glycosyltransferase"  
FT /note= "encodes AAB82210"  
FT complement(10483..11424)  
FT /*tag= k  
FT /gene= "megBIV"  
FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
FT /note= "encodes AAB82211"  
FT 12181..22821  
FT /*tag= l  
FT /gene= "megA"  
FT /product= "megalomycin 6-deoxyerithronolide B synthase 1"  
FT /note= "polyketide synthase, encodes AAB82212"  
FT 12505..13470  
FT /*tag= m  
FT /gene= "megA"  
FT /function= "AT-L"  
FT 13576..13791  
FT /*tag= n  
FT /gene= "megA"  
FT /function= "ACP-L"  
FT 13849..15126  
FT /*tag= o  
FT /gene= "megA"  
FT /function= "KS1"  
FT 15427..16476  
FT /*tag= p  
FT misc_feature  
FT misc_feature  
FT misc_feature
```

```
FT /gene= "mega"
FT /function= "AT1"
FT misc_feature
FT /tag= q
FT /gene= "mega"
FT /function= "KR1"
FT misc_feature
FT /tag= x
FT /gene= "mega"
FT /function= "ACp1"
FT misc_feature
FT /tag= b
FT /gene= "mega"
FT /function= "KS2"
FT misc_feature
FT /tag= c
FT /gene= "mega"
FT /function= "AT2"
FT misc_feature
FT /tag= u
FT /gene= "mega"
FT /function= "KR2"
FT misc_feature
FT /tag= v
FT /gene= "mega"
FT /function= "ACp2"
FT CDS
FT /tag= w
FT /gene= "megaili"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT /note= "polyketide synthase, encodes AAB8213"
FT misc_feature
FT /tag= x
FT /gene= "megaili"
FT /function= "KS3"
FT misc_feature
FT /tag= ai
FT /gene= "megaili"
FT /function= "KS5"
FT misc_feature
FT /tag= y
FT /gene= "megaili"
FT /function= "AT3"
FT misc_feature
FT /tag= aj
FT /gene= "megaili"
FT /function= "AT5"
FT misc_feature
FT /tag= z
FT /gene= "megaili"
FT /function= "KR3 (inactive)"
FT misc_feature
FT /tag= ak
FT /gene= "megaili"
FT /function= "KR5"
FT misc_feature
FT /tag= aa
FT /gene= "megaili"
FT /function= "ACp3"
FT misc_feature
FT /tag= ai
FT /gene= "megaili"
FT /function= "ACp5"
FT misc_feature
FT /tag= ab
FT /gene= "megaili"
FT /function= "KS4"
FT misc_feature
FT /tag= am
FT /gene= "megaili"
FT /function= "KS6"
FT misc_feature
FT /tag= ac
```

```
FT /gene= "megaili"
FT /function= "AT4"
FT misc_feature
FT /tag= an
FT /gene= "megaili"
FT /function= "AT6"
FT misc_feature
FT /tag= ad
FT /gene= "megaili"
FT /function= "DH4"
FT misc_feature
FT /tag= ao
FT /gene= "megaili"
FT /function= "KR6"
FT misc_feature
FT /tag= ae
FT /gene= "megaili"
FT /function= "ER4"
FT misc_feature
FT /tag= ap
FT /gene= "megaili"
FT /function= "ACPe"
FT misc_feature
FT /tag= af
FT /gene= "megaili"
FT /function= "KR4"
FT misc_feature
FT /tag= aq
FT /gene= "megaili"
FT /function= "TB"
FT misc_feature
FT /tag= ag
FT /gene= "megaili"
FT /function= "ACp4"
FT CDS
FT /tag= ah
FT /gene= "megaili"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 3"
```

```
Query Match 100.0%; Score 600; DB 4; Length 47981;
Best Local Similarity 100.0%; Pred. No. 3,6e-85;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCTGCGCGAGGTACACCCGGGGGTGAGGTGCACTGCGCTACCGCAGTGGTG 60
DB 36307 CCGCGCTGCGCGAGGTACACCCGGGGGTGAGGTGCACTGCGCTACCGCAGTGGTG 36366
QY 61 AGGAGCGCCGGTGCAGCTGCGGCTTACCCGTTCAACGACGAACTTGGCTCCCGG 120
DB 36367 AGGAGCGCCGGTGCAGCTGCGGCTTACCCGTTCAACGACGAACTTGGCTCCCGG 36426
QY 121 TCCCCCTGGAGCGGCTCCCGACACCGGAGCAGAGTGAGGCTTACGACTGCTGCGACC 180
DB 36427 TCCCCCTGGAGCGGCTCCCGACACCGGAGCAGAGTGAGGCTTACGACTGCTGCGACC 36486
QY 181 CCGTGCAGCTCGGCGGCTCTCCCTGCGCGGACGAGTCTCTGTGTGACCGGACGCGCAG 240
DB 36487 CCGTGCAGCTCGGCGGCTCTCCCTGCGCGGACGAGTCTCTGTGTGACCGGACGCGCAG 36546
QY 241 TACCCCGGCTGAGCAGAGTGATCCGCGACGAGCTGGAACAGCGCGGGGCGACCGGTG 300
DB 36547 TACCCCGGCTGAGCAGAGTGATCCGCGACGAGCTGGAACAGCGCGGGGCGACCGGTG 36606
QY 301 TGTGTGACCGCGAGTGCGCGCGCCCGGATCGGCGCGCACTCGAAGCGCTGACGCGCA 360
DB 36607 TGTGTGACCGCGAGTGCGCGCGCCCGGATCGGCGCGCACTCGAAGCGCTGACGCGCA 36666
QY 361 CGCCCTGTCCACTGTGTGTCTCTGCTGCGCGCTGCGCGAGGCGGTGCTGTGACGAGCC 420
DB 36667 CGCCCTGTCCACTGTGTGTCTCTGCTGCGCGCTGCGCGAGGCGGTGCTGTGACGAGCC 36726
QY 421 CCAGCCTGACACCTCGGCTGTGTGTCAGGCGCTGCGCGACCGGAGTCGAGTCCGCC 480
```


PR 26-SEP-2002; 2002US-0414085P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Santi DV, Reid RC, Kodumal SJ, Jayaraj S;
XX WPI; 2004-316102/29.
XX
XX New synthetic genes encoding polyketide synthases, useful for
XX facilitating production of therapeutic polyketide compounds, or in human
XX and veterinary medicine, pharmacology, agriculture and molecular biology.
XX
XX Example 7, Page 110-112; 172pp; English.
XX
XX The present invention relates to a synthetic gene encoding a polypeptide
XX segment that corresponds to a reference polypeptide segment encoded by a
XX naturally occurring gene. In particular, the gene is a polyketide
XX synthase coding sequence. The composition, the gene and methods are useful for
XX producing synthetic genes or libraries of such genes, and for
XX manipulating and characterizing genes and corresponding encoded
XX polypeptides. The synthetic genes may be used for facilitating production
XX of therapeutic or commercially important polyketide compounds. These may
XX be used in human and veterinary medicine, pharmacology, agriculture and
XX molecular biology. The present sequence is a synthetic gene used in the
XX exemplification of the invention.
XX
SQ Sequence 9510 BP; 1587 A; 2635 C; 3236 G; 2052 T; 0 U; 0 Other;

Query Match 31.5%; Score 189; DB 12; Length 9510;
Best Local Similarity 59.5%; Pred. No. 4.1e-21;
Matches 359; Conservative 0; Mismatches 235; Indels 9; Gaps 2;

QY 1 CCGCGCTCCGCGAGCGGTACACCGCGGGGTGAGGTGCACTGCGCCGACCGAGTGGTGG 60
DB 2582 CCGCATTTAGCGAAGCCTATGCCAGGCGGTGAGGTGCACTGCTTTGGCGG 2641
QY 61 AGGAGCGCCCGGTGCACTGCGGTACCGGTTCACAGACAGAACTTGTGGTCCCGG 120
DB 2642 ATCCCGGCCAGTGGAAATTACAGTGTATCCGTTTCAGCCGTACCGTTACTGGCTGAGA 2701
QY 121 TCC---CCCTGGGCGGGGTCCCGACACCGGCGACGAGTGGCGTTACGCTCGCTGGC 177
DB 2702 TTCGACAGGTGGGCGGGCTCGTACAGAAATGATGATGCGTTACAGGTGGTTGGC 2761
QY 178 ACCCGCGACCTCGGGCGGTCTCCCTGCGCGGAGGCGGTCTGGTGTGACCGGAGCGG 237
DB 2762 GTGAGCGGAAATGGAGTCTGCGTCTCGCGGTGCGGTGCTGTGTAACCGGCCCGG 2821
QY 238 CAGTACCCCGCGCTGAGCGGAGTGGTCCGCGACCGCTGGAACAGCGGGGCGACCG 297
DB 2822 GTGTACCATCTGAGCTGTCCGATGCCATCCGCTCAAGGGCTGAGACAGTCCGGGGCGA 2881
QY 298 TCGTGTGTGTCACCGCGCAGTCCGCGCGCCGAGTCCGCGCGCACTGACGCGGTGACG 357
DB 2882 TTTTGACATGCAAGTCAAGAAAGCGTTCCAGATCCGACGCGGCTTGGAAGTGTGATA 2941
QY 358 GCAACCGCCCTGTGCACTGTGTCTCTCTGCTCGCGCTCGCGGAGGGCGGTGTGTGACG 417
DB 2942 CTGATCGCTGACGACCGGTATGCTGTGTAACCGGTG-----ATGGGAGGCTGTG 2995
QY 418 ACCCGACCTGACACCGCTCGGTGTGTCAGGCGCTCGCGACCGCGGAGTGCAGCTGC 477
DB 2996 ATCCGAGTCTGATGCTGTGCTTTGGTGAAGGCGCTTGGTGTCTGCTGGCGTGAAGCAC 3055
QY 478 CCGTGTGGTGTGACACGAGGACCGCGCGCTGACCGCTGAGACGAGTGCATCCG 537
DB 3056 CGCTGTGGGTCTGACCGCGTAATGCTGTCAAGGTGCTGATGTGAGTGTGATGCTG 3115
QY 538 CCGAGCGCATGTGCTGTGGTGTGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 597
DB 3116 CCGAAGCGATGTGTGGGCGGT 3175
QY 598 GTG 600

DB 3176 GCG 3178

RESULT 5
AAD55818
ID AAD55818 standard; DNA; 4725 BP.
XX
XX AAD55818;
AC
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
XX Micromonospora carbonacea polyketide synthase (PKS) type I gene #8.
DE
XX Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;
KW gene; de.
XX
XX Micromonospora carbonacea.
OS
FH Key Location/Qualifiers
FT CDS 1..4725
FT /tag= a
FT /product= "polyketide synthase"
XX
XX CA2391131-A1.
XX
XX 19-NOV-2002.
XX
XX 26-JUL-2002; 2002CA-02391131.
XX
XX 26-JUL-2001; 2001US-0307629P.
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX Yang X, Staffa A, Farnet CM;
XX WPI; 2003-343556/33.
XX P-PSDB; AAE37002.
XX
XX Novel isolated polypeptide involved in biosynthesis of macrolides by
XX PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
XX PT preferably for biosynthesis of rosamycin.
XX
XX Claim 1; Page 164-166; 206pp; English.
XX
XX The invention relates to genes and proteins involved in the biosynthesis
XX of macrolides by microorganisms. In particular it relates to the nucleic
XX acids forming the biosynthetic locus for rosamycin (a 16-member
XX macrolide antibiotic) from Micromonospora carbonacea. The invention is
XX useful for the biosynthesis of macrolides by microorganisms. It allows
XX direct manipulation of the proteins involved in the biosynthesis of
XX rosamycin. It is useful to catalyse certain biochemical reactions, in
XX vitro or in vivo, to direct or enhance the synthesis or modification of a
XX polyketide, polyketide substrate or its precursor. The present sequence
XX is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
XX OCT-2003 to standardise OS field)
XX
SQ Sequence 4725 BP; 482 A; 1782 C; 1885 G; 576 T; 0 U; 0 Other;

Query Match 20.7%; Score 124; DB 10; Length 4725;
Best Local Similarity 55.0%; Pred. No. 5.9e-11;
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;

QY 1 CCGCGCTCCGCGAGCGTACACCGGGGGGTGAGGTGACT-----GACGTACCGCAG 54
DB 2642 CTTCCGCGGCGCAGAGGCTTACACCGGGGGGTGAGGTGACTGTGCTCCGGGCTTGGCG 2701
QY 55 TGGGTGAGGAGGACCGCGGTGCACTGCGGTCTTACCGGTTCACAGACAGAACTTGTGGC 114
DB 2702 GGGCGCGCGCGCTCGCGGTGACCTGCGCACGTGACGCTTCCAGCGGCAACGCTACTGGC 2761

QY 115 TCCCGATCCCCCTGAGCGGAGTCCCGACA---CCGCGACGAGTGGCGTTACCACTCG 171
Db 2762 TGAAGGCCCGCGCGAGCGGAGACTCCGAGCCCGGAGACGGCTGGCGCTACCGGGTCG 2821
QY 172 CTTGGACCCCGTGCACCTTGAGCGGCGTCTCTCTGCGGAGCGGAGTCTTGATGACCG 231
Db 2822 GCTGGCGCGGCTTTCACCGGACCGGCGCGCGCGCGGAGCCGGTGGCTCTGTGACCG 2881
QY 232 GAGCGGACGATACCCCGGCGCTGACCGGACGATGTCGCGGACGCGCTGGAACAGCGCGGG 291
Db 2882 GCCCGAGACGAGCCCGGAGAGTGTGAGAGCGGAGCGGCTACACGCGCGGGGCG 2941
QY 292 CGACCGTGTGTGTGACCGCGGACGTCG-----CGGCGCGGATCGCGCG 339
Db 2942 CCGAGGTCGCGCTGTGATCGGTGAGCGGACGACGACCGAGCCGGCGCGGCGCGGT 3001
QY 340 CACTCGACCGCGTGCAGCGGACCGGCGCTGTTCACGTGTGTCTCTGCTCG----- 390
Db 3002 TGCTCACCGCGCGGTCGCGGCGGCGGAGACCGCGGTCGTCTGCTCGGACCGATC 3061
QY 391 -----GGCTGCGCGGAGGCGGTCGTGACGACCCGAGCTGGAACCGCTGCGGTTGG 444
Db 3062 GTGCGCGGACCCGACCGACCGGCGGTCGTGCGGCGCGCGGAGCTTGCTGCTGA 3121
QY 445 TCCAGCGCTCTGCGCGACGCGGAGTGCAGTCCTCCCTGTGTGCTGTGACGAGGACCGCG 504
Db 3122 CCGAGGCGGTGCGCGAGCGGCTGCGCGCGCGCGGCTGTGCTGTGACCGCGGCGCG 3181
QY 505 CCGCGGTGACCGTGCAGAGACGATCGATCCGCGCGGAGCGGATGTGATGAGGCTCGGCG 564
Db 3182 TCTCCGTGCGGCGCGGAGACCGCGGACGAGCGGCGGAGCGGAGGCTGTGCGGCG 3241
QY 565 GGGTGTGCGGAGTGTGCGGCGCGCGCGCGGCGGCGG 600
Db 3242 GGGTGTGCGGCGGCTGCACTGCGCGCGGAGTGTGCGGCG 3277

RESULT 6
AAD55810
ID AAD55810 standard; DNA; 60196 BP.
XX
XX
AC AAD55810;
DT 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DE Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.
XX
KM Macroliide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KM gene; ds.
XX
XX
OS Micromonospora carbonacea.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1683
FT FT /*tag= a
FT FT /product= "Polyketide synthase #1"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT CDS complement(1728..2522)
FT FT /*tag= b
FT FT /product= "Polyketide synthase #2"
FT FT complement(2629..3861)
FT FT /*tag= c
FT FT /product= "Polyketide synthase #3"
FT FT 4365..5573
FT FT /*tag= d
FT FT /product= "Polyketide synthase #4"
FT FT 5702..19117
FT FT /*tag= e
FT FT /product= "Polyketide synthase #5"
FT FT 19144..24921
FT CDS /*tag= f

FT FT /product= "Polyketide synthase #8"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT CDS 2493..36230
FT FT /*tag= g
FT FT /product= "Polyketide synthase #7"
FT FT 36292..41016
FT FT /*tag= h
FT FT /product= "Polyketide synthase #8"
FT FT 41049..46403
FT FT /*tag= i
FT FT /product= "Polyketide synthase #9"
FT FT 46400..47794
FT FT /*tag= j
FT FT /product= "Polyketide synthase #10"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT CDS 47794..49083
FT FT /*tag= k
FT FT /product= "Polyketide synthase #11"
FT FT 49092..49814
FT FT /*tag= l
FT FT /product= "Polyketide synthase #12"
FT FT 49868..51226
FT FT /*tag= m
FT FT /product= "Polyketide synthase #13"
FT FT /transl_except= (pos:51221..51223, aa:Yaa)
FT FT /note= "Yaa corresponds to amino acids from position 452-1811"
FT CDS 51506..53416
FT FT /*tag= n
FT FT /product= "Polyketide synthase #14"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT CDS complement(54569..53358)
FT FT /*tag= o
FT FT /product= "Polyketide synthase #15"
FT FT 54897..56342
FT FT /*tag= p
FT FT /product= "Polyketide synthase #16"
FT FT 56408..57634
FT FT /*tag= q
FT FT /product= "Polyketide synthase #17"
FT FT 57657..59123
FT FT /*tag= r
FT FT /product= "Polyketide synthase #18"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT CDS 59363..60196
FT FT /*tag= s
FT FT /product= "Polyketide synthase #19"
XX
XX
XX CA2391131-A1.
XX
XX
XX 19-NOV-2002.
XX
XX
XX 26-JUL-2002; 2002CA-02391131.
XX PF
XX 26-JUL-2001; 2001US-0307629P.
XX
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX
XX Yang X, Staffa A, Farnet CM;
XX
XX
XX WPI; 2003-343556/33.
XX
XX P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,
XX DR AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,
XX DR AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.
XX
XX
XX Novel isolated polypeptide involved in biosynthesis of macroliides by
XX PT microorganisms, useful for biosynthesis of macroliides by microorganisms,
XX PT preferably for biosynthesis of rosaramicin.

PS Example 2, Page 59-94; 206pp; English.

XX The invention relates to genes and proteins involved in the biosynthesis
CC of macrolides by microorganisms. In particular it relates to the nucleic
CC acids forming the biosynthetic locus for roxamycin (a 16-member
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC useful for the biosynthesis of macrolides by microorganisms. It allows
CC direct manipulation of macrolides and related chemical structures by
CC chemical engineering of the proteins involved in the biosynthesis of
CC roxamycin. It is useful to catalyse certain biochemical reactions, in
CC vitro or in vivo, to direct or enhance the synthesis or modification of a
CC polypeptide, polypeptide substrate or its precursor. The present sequence
CC is M. carbonacea polypeptide synthase (PKS) type I gene cluster. (Updated
CC on 27-OCT-2003 to standardise OS field)

XX Sequence 60196 BP; 6934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;

Query Match 20.7%; Score 124; DB 10; Length 60196;
Best Local Similarity 55.0%; Pred. No. 3.9e-11;

Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;

QY 1 CCGGCGTCCGCGGAGGCTTACACCGGGGGTGAAGTCACT-----GGCTAACCGCAG 54
DB 38933 CCTCGGCGGCGGAGGCTTACACCGGGGGTGAAGTCACTCGGCGTCTGCGCG 38992
QY 55 TGGGTAGAGGAGACCGCGGCTGACCTGCGGCTTACCGGTTCAACGACAGACTTCTGGC 114
DB 38993 GGGCGCGCGCGCGCTGCGGCTGACCTGCGGCTTACCGGTTCAACGAGCGTACTGGC 39052
QY 115 TCCCGGTCCCGCTGGGCGGCGGCTCCGACA---CCGCGACGAGTGGCGTTACAGCTCG 171
DB 39053 TGGAGCCCGCGCGGCGGAGCTCCGCGCGCGCGCGGCGGAGCGGCTGCGGCTCG 39112
QY 172 CCTGGACACCGCGCTGACCTGCGGCGGCTCTCTCGCGCGGAGCGGCTGCTGGTGAACG 231
DB 39113 GCTGGCGCGCGCTTACGCGGAGCGGCGCGCGCGCGCGCGGAGCGGCTGCTGGTGAACG 39172
QY 232 GAGCGGAGTACACCGCGCGCTTGAACGAGTGGTGGCGGCGGCGCTGGAACAGCGGCGG 291
DB 39173 GCGCGGAGACGCGCGGAGAGCTGTGAGGCGGCTGCGGCGGCTTACCGCGCGGCGGCG 39232
QY 292 CGACCGCTGCTGTGTGACACCGCGCAGTGC-----CGCGCGCGGAGTGGCGCG 339
DB 39233 CGAGGCTGGCGCTGTGAGCGGCTGAGCGGAGCGGAGCGGCGGCTTACCGCGCGGCGG 39292
QY 340 CACTGACACCGCGCTGACCGGACCGCGCTGTGCTGCTGTGCTGCTGCTGCTGCTGCTG 390
DB 39293 TGGTCAACCGCGCGGCTGTGCGGCGGCGGAGCGGCGGCTGTGCTGCTGCGACCGATC 39352
QY 391 -----CGCTCGCGGAGGCGGCTGTGCGGAGCGGCGGCGGAGCGGCGGCTGCG 444
DB 39353 GTGCGCGGAGCGGCGGAGCGGCGGCGGCTGTGCGGCGGCGGCGGAGCGGCTGCTGA 39412
QY 445 TCCAGGCGCTGCGCGGAGCGGCGGAGTGAAGTCCCGCTGTGCTGTGAGCAGGAGCGCG 504
DB 39413 CCCAGGCGGCTGCGCGGAGCGGCGGCGGCGGCGGCTGTGCGGCTGTGCGGCGGCG 39472
QY 505 CGGCGCTGACCGCTGCGGAGCGGAGTGAAGTCCCGCGGAGCGGCGGAGTGTGCGGCTGCG 564
DB 39473 TCTCCCTGCGGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGTGTGCGGCTGCG 39532
QY 565 GGGTGGTGGGCGGCTGAGTCCCGCGGCGGCGGCGG 600
DB 39533 GGGTGGGCGGCGGCTGAGTCCCGCGGAGCGGCGGCG 39568

RESULT 7
AADI17186
ID AADI17186 standard; DNA; 125401 BP.
AC AADI17186;
XX
XX
DT 29-NOV-2001 (first entry)

XX Streptomycetes noursei nystatin PKS gene cluster DNA.
DE
XX Polypeptide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; ds.
XX
OS Streptomycetes noursei.

XX Key Location/Qualifiers
FH CDS 6337..34771
FT /tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /tag= d
FT /product= "NysL protein"
FT complement(58786..58980)
FT /tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement(59045..60241)
FT /tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement(60238..61296)
FT /tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /tag= h
FT /product= "NysR4 (long) protein"

PN MO200159126-A2.

PD 16-ANG-2001.

PP 08-FEB-2001; 2001WC-GB000509.

PR 08-FEB-2000; 2000GB-00002840.

PR 10-APR-2000; 2000GB-00008786.

PR 14-APR-2000; 2000GB-00009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.
PA (SNTF) SINTER STIETELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINY-) SINVENT AS.
PA (DZIE/) DZIELEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FUAE/) FUAEV E.
PA (BRAU/) BRAUTASSET T.
PA (STRO/) STROM A R.
PA (VALU/) VALUA S.

XX Zotchev SB, Sekurova ON, Fjaervik R, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
PI
XX

DR WPI, 2001-557614/62.
DR F-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.

XX New nystatin polypeptide synthase polynucleotides and polypeptides, useful
PT as antibiotics and antifungals.
PT
XX

PS Claim 1, Page 188-254; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polypeptide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin. The

CC nystatin PKS is useful as antifungal antibiotic. The present sequence is
CC a *Streptomyces noursei* nystatin PKS gene cluster DNA
XX
S0 Sequence 125401 BP, 15664 A, 43692 C, 42871 G, 17174 T, 0 U, 0 Other;

Query Match	17.2%	Score 103.2;	DB 4;	Length 125401;
-------------	-------	--------------	-------	----------------

Matches 291; Conservative 0; Mismatches 238; Indels 15; Gaps 3;

OY	72	GTGCACCTGCAGCGGCTACCCGGTTCCAAACGACAAACATTTCCGGATCTCCGGTCCCTCCGGACC	131
Db	43840	GGGCAACCAAGCTCCGGCGCCCGGCTGTCGCCGCACTGTCCGCTTGGCGACACCCGGGCGCCGACAC	43899
OY	132	CGGGTCCCGGACACCGGCGGACGAGTGGCGTTTACCAAGCTCGCCCTGACACCCCGTGCACCTTC	191
Db	43900	AACATGGGCGACACCAACGCGCTGGCGGACACCGGAGAGATTGGGAAACCGCTGTCCGTCGCGCG	43959
OY	192	GGGCGGTCCTCCCTGGCGCGGACCGGGTCTGTGGTGTACCGGAGCG---GCAGTACCCCG	248
Db	43960	AAGCGGACACACCGCGCGCGTCTCTGTGTGTGTCCCGCGCGCGCGACACACCGACCCCTTGG	44019
OY	249	GCCTGGACGGAACGTGGTCCCGCGACCGCGCTGGAAACAGCGCGGGGCGACCCGATGGTTTGTGC	308
Db	44020	GTTCGCCCAAGTCCGTCCGCGCGCTTCGGCGCGGAGACGCCCGCGGGTGGACGTTCCCGGCGAC	44079
OY	309	ACCAGGAGATGCAGCGCGCCGGATTCGAGCGACGCACTGCAGCGCGCTGCACCGGACACCGCGCTTG	368
Db	44080	GGCACACCAACGGGGCGCGCGTTCGCGCGCGCTGTCAACGAGAGCGGCGGACACACACACCGCGCG	44139
OY	369	TCCACTGGATCTCTTGTGCTTCGCGCTCGCGAGGGCGGGTCTGTGCACGA-----C	419
Db	44140	ACCACCGGTGCTCCCTGCTCGCGCTGCAGACGACGACGCGACGACCGCGGTACCGGCG	44199
OY	420	CCGAGCGCTGGACACCTCCGCGGTTGGTTCAGAGCGCTGGGCGACGCGGAGATTCGACGTCGCCCC	479
Db	44200	GGCACACACCGGACACCGCGCGCGCTGTTCAGAGGACCTTCGCGCACACCGGACGCGCCGCGCGCG	44259
OY	480	CTGTGGCTGTGGACCAAGGAGACG---GCGCGCGTGCACGCTCGGAGACGACGTCGATCCG	536
Db	44260	CTGTGGCGCCCTGCACCCCGGGGGCGGGTGGCGCGCGCTCCCGGACGAGCGCGACCGCGCCC	44319
OY	537	GCCACGCGCATGTCCGATGGCTCGGCGCGGGTGGGCGTGGAGTCCCCCGCGCGGGTGG	596
Db	44320	GCCACGCGCGCGCTGTGGGGCGCTCGGCGCGAGTGCGCGCCTTCGAACTTCGCGCGCACACTGG	44379
OY	597	GATG 600	
Db	44380	GGCG 44383	

RESULT 8

ID AAD55817 standard; DNA; 11238 BP.

AC AAD55817;

DT	27-OCT-2003	(revised)
DT	07-AUG-2003	(first entry)

DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.

KM Macrolide; rosamycin; polyketide synthase, PKS; enzyme gene, ds.

OS *Micromonospora carbonacea*.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	/tag= a
----	---------

FT /note= "CDS does not include start codon"

XX

PN CA2391131-A1.

PD 19-NOV-2002

PF 26-JUL-2002; 2002CA-02391131.

PR 26-JUL-2001; 2001US-0307629P.

PA (ECOP-) ECOPIA BIOSCIENCES INC.

PI Yang X, Stafra A, Farnet CM;

DR WPI; 2003-343556/33.

[illegible]

PT microorganisms, useful for biosynthesis of macrolides by microorganisms
PT preferably for biosynthesis of rosamycin.

PS Claim 1; Page 151-158; 206pp; English.

The invention relates to genes and proteins involved in the biosynthesis of macrolides by microorganisms, in particular it relates to the nucleic acids forming the biosynthetic locus for rosamycin (a 16-member macrolide antibiotic) from *Micromonospora carboxa*. The invention is useful for the biosynthesis of macrolides by microorganisms. It allows direct manipulation of the macrolides and related chemical structures by chemical engineering of the proteins involved in the biosynthesis of rosamycin. It is useful to catalyse certain biochemical reactions, in vitro or in vivo, to direct or enhance the synthesis or modification of polyketide, polyketide substrate or its precursor. The present sequence is *M. carboxa* polyketide synthase (PKS) type I gene. (Updated on 27-Oct-2003 to standardise OS field)

Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;

Query Match	Score	DB	Length
16.74	100.2	10	11238

Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1.

QY	143	TACCGGACACAGATGAGCGCTTACACAGACTGAGCTGAGCACCCTGACACTCTGGGCGAGTCTCTC	202
Db	2949	CACCTGGAAAACCGTTGCGGACCGACGACGACCCCGAACCCCTCACAACAACCGCTGGCTGTGAT	3008
QY	203	CCYGGCCGGAACGAGGTCTGTGTGTGTGACCGGAGCGGACAGTACCCCGCGCTTGTGAAGACGT	262
Db	3009	GGTGCCCCGAGGAGCGCTGAGCCGGATGTACGAGGATGTGTGGCGGCTGTACCGGACAGCGCTGGCGGC	3068
QY	263	GGTCCGCGACACGGCCTTGGAACAGACCGCGGGGCGAACCGTGTGTGTGTGACACCGGCAAGTCCGC	322
Db	3069	GGTGAGCGCTCTCGTGTGAATGTCTGACCGTGTGGCACCGCGCGAACCGGAGCGGCGCTTGTGCGC	3128
QY	323	CGCCCGGATTCGGCGCGCGGACCTGACAGCGCGGTGACAGGCAACCGCCCTGTTCACATGTGTGTTC	382
Db	3129	GCTCTGTGACTCCGCGCGCGCCCGCGGACCCGGAAGCGCGCGCGCCCGCGCGGCGGTGTCTC	3188
QY	383	TCGTCTCGGCTCTCGCGAGGCGGAGTCTGTGTGACGACCCCAAGCTG-----	428
Db	3189	CTGTGTGCGGCTCTCGCGAGGCGCGGACCGCGCGCACCCCGGACCGTACACGCTTCGCGTTGAC	3248
QY	425	-GACACCCCTCGGTGTGTGTGACAGGCGCTCGGCGGACCGGAAATGCAAGTCCCCCTGTGGCT	487
Db	3249	CGCTCTCGGCTCTGTGATTCAGAGCAATTGACCGGACCGCGGAGAACAGAGCGCCGCTTGTGGC	3308
QY	488	GGTGACACAGGGAACGCGCGCGCGCGCTGTGACCGATCGGAGACGACGTCGATTCGATTCGAGCCCAAGCCAT	547
Db	3309	GGTACACCGGAGGAGCGCGTGTGCGGTCTCTCCGCGGAGGATGCGGAGCGCGGAGCAGGCCCA	3368
QY	548	GGTCCGTGTGCTCTCGGCGCGGATGTGTGCGCGTGAATCTCCCGCGCGGATGTGGGTG	600
Db	3369	GGTGTGTGGGCTCTCGCGCGGATGTGCGGCGCTCGAATGTGCGGACCGATGTGGGGG	3421

```

RESULT 9
AD251725/c
XX AD251725 standard; DNA; 138203 BP.
XX
XX AD251725;
XX
XX 14-UTL-2005 (first entry)
XX
XX FR-008 polyketide gene cluster, SEQ ID 1.
XX
XX Antifungal; antiparasitic; polyketide; gene; ds.
XX
XX Streptomyces sp.; FR-008.
XX
XX Key
XX Location/Qualifiers
XX FT complement(574..1950)
XX FT /tag= a
XX FT /product= "FAD-dependent monooxygenase, FscO, SEQ ID 20"
XX FT complement(2264..3037)
XX FT /tag= b
XX FT /product= "ADC lyase, PabC, SEQ ID 22"
XX FT complement(3150..3818)
XX FT /tag= c
XX FT /product= "Regulator protein, FscRI, SEQ ID 10"
XX FT complement(4377..7205)
XX FT /tag= d
XX FT /product= "Regulator protein, FscRII, SEQ ID 11"
XX FT complement(7210..10320)
XX FT /tag= e
XX FT /product= "Regulator protein, FscRII, SEQ ID 12"
XX FT complement(10298..13315)
XX FT /tag= f
XX FT /product= "Regulator protein, FscRIV, SEQ ID 13"
XX FT 13522..14898
XX FT /tag= g
XX FT /product= "Glycosyltransferase, FscMI, SEQ ID 17"
XX FT 14953..16011
XX FT /tag= h
XX FT /product= "GDP-ketosugar aminotransferase, FscMII, SEQ ID 18"
XX FT 16106..17287
XX FT /tag= i
XX FT /product= "Cytochrome P-450 monooxygenase, FscP, SEQ ID 14"
XX FT 17334..17528
XX FT /tag= j
XX FT /product= "Ferredoxin protein, FscFB, SEQ ID 15"
XX FT 17556..18413
XX FT /tag= k
XX FT /product= "Thioesterase, FscTB, SEQ ID 16"
XX FT 18610..20781
XX FT /tag= l
XX FT /product= "ADC synthase, PabAB, SEQ ID 21"
XX FT 20927..26158
XX FT /tag= m
XX FT /product= "PKS FscA, SEQ ID 2"
XX FT 26333..27340
XX FT /tag= n
XX FT /product= "ABC transporter, FscTI, SEQ ID 8"
XX FT 27561..28280
XX FT /tag= o
XX FT /product= "ABC transporter, FscTII, SEQ ID 9"
XX FT complement(28983..60860)
XX FT /tag= p
XX FT /product= "PKS FscC, SEQ ID 3"
XX FT complement(60962..77587)
XX FT /tag= q
XX FT /product= "PKS FscB, SEQ ID 4"
XX FT complement(77983..84132)
XX FT /tag= r
XX FT /product= "PKS FscF, SEQ ID 5"
XX FT complement(84170..107485)
XX FT /tag= s

```

```

FT FT /product= "PKS FscE, SEQ ID 6"
FT FT complement(107496..136148)
FT FT /tag= t
FT FT /product= "PKS FscD, SEQ ID 7"
FT FT complement(136558..137766)
FT FT /tag= u
FT FT /product= "GDP-mannose-4,6-dehydratase, FscMII, SEQ ID 19"
XX
XX US2005089884-A1.
XX
XX 28-APR-2005.
XX
XX 06-APR-2004; 2004US-00819386.
XX
XX 23-OCT-2003; 2003KR-00074035.
XX
XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
XX (UYSH-) UNIV SHANGHAI JIAOTONG.
XX
XX Lee SY, Deng Z, Chen S, Jeong KJ, Zhou X;
XX
XX WPI; 2005-322033/33.
XX
XX P-PSDB; AD251726, AD251727, AD251728, AD251729, AD251730, AD251731,
XX AD251732, AD251733, AD251735, AD251736, AD251737, AD251738,
XX AD251739, AD251740, AD251741, AD251742, AD251743, AD251744, AD251745,
XX AD251746.
XX
XX New gene cluster for the biosynthesis of FR-008 polyketides, useful for
XX increasing the productivity of FR-008 polyketides or to produce new FR-
XX 008 polyketide variants.
XX
XX Claim 1; SEQ ID NO 1; 22pp; English.
XX
XX The present invention relates to a gene cluster (AD251725) for the
XX biosynthesis of FR-008 polyketides derived from Streptomyces sp. FR-008.
XX FR-008 polyketide, a heptene macrolide having aglycone containing 4-
XX aminocetophenone, has antifungal activity and also high toxicity against
XX mosquito larvae, and so is highly useful in agricultural and medical
XX fields. The gene cluster comprises the following genes: fscA, fscC, fscB,
XX fscF, fscB and fscD, which code for modular polyketide synthase (PKS);
XX fscRI and fscRII, which code for ABC transporter proteins; fscRI, fscRII,
XX fscRIII and fscRIV genes which code for regulator proteins; fscP which
XX codes for cytochrome P450 monooxygenase; fscTB which codes for ferredoxin
XX protein; fscTB which codes for thioesterase; fscMI which codes for
XX glycosyltransferase; fscMII which codes for GDP-mannose-4,6-dehydratase;
XX aminotransferase; fscMIII which codes for GDP-mannose-4,6-dehydratase;
XX fscO which codes for FAD-dependent monooxygenase; PabAB which codes for 4
XX -amino-4-deoxychorismate (ADC) synthase and PabC which codes for ADC
XX lyase. The genes of the gene cluster of the invention can be used to
XX develop recombinant microorganisms capable of producing FR-008
XX polyketides, also to increase the productivity of the existing FR-008
XX polyketides or to produce new FR-008 polyketide variants, by its
XX modification.
XX
XX Sequence 138203 BP; 18153 A; 48202 C; 55342 G; 16506 T; 0 U; 0 Other;
XX
XX Query Match 16.7%; Score 100.2; DB 14; Length 138203;
XX Best Local Similarity 53.7%; Pred. No. 1.7e-07;
XX Matches 341; Conservative 0; Mismatches 258; Indels 36; Gaps 5;
XX
XX 2 CCGGCTCCGCGAGGCTACACCCGGGGGTGAGTGCACCTGACCGGCTGAGTGA 61
XX 92504 CCGGCTTCAGCGCGGCTTCTTGACCGCCGCTCAGCGGAGACCTGACCGCTCCGCC 92445
XX
XX 62 GGAAGCGCCCGGTGACCTGCGGCTTACCGTTCACGACGAACTTCTGCGTCCGGGT 121
XX 92444 CGACCTGCGCGTGGACACCGAGGCGCTGCGGCGGTATCTCCGGCCCTTCCACCTCGGCG 92385
XX
XX 122 CCCCCCTGGGCGGGTCCCGACACCGGCGAGGAGTGGCGTTACCAAGCTGCGCTGACA--C 179
XX 92384 GGGCGGCGCGCGGAGACGAGGCGCATGGTGACTCCAAACGACCAAGAGACGTGAGAGCC 92325

```


CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or ~
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narboxmycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macroclide biosynthetic gene clusters are useful for the production
CC of biologically active macroclides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narboxmycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer syntheses or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439,
XX as given in figure 31. (Updated on 15-SEP-2003 to standardise OS field)

Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 0 U; 2 Other;

derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or
Streptomyces antibioticus. The invention also relates to a macrolide
 biosynthetic gene cluster, or fragments thereof. The macrolide
 biosynthetic gene cluster encodes proteins which synthesise methymycin,
 pikromycin, neomethymycin, narboxycin or a combination of these
 compounds. Recombinant or augmented cells comprising the deoxamine
 and/or macrolide biosynthetic gene clusters are useful for the production
 of biologically active macroclides. The macrolide biosynthetic proteins
 are useful for synthesis of methymycin, pikromycin, neomethymycin and
 narboxycin. The alternative termination of polyketide synthesis may be
 useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 monomers. The compounds produced by the recombinant host cells are useful
 as biopolymers, e.g., in packaging or biomedical applications, to
 engineer PHA monomer synthases or to prepare biologically active agents,
 such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 chronic obstructive pulmonary disease as well as other diseases involving
 respiratory inflammation, cholesterol-lowering agents or macroclide-based
 antibiotics which are active against a variety of organisms, e.g.,
 bacteria, including multi-drug resistant pneumococci and other
 respiratory pathogens, as well as viral parasitic pathogens, or as crop
 protection agents (e.g., fungicides or insecticides) via expression of
 polyketides in plants. The present sequence represents the macroclide
 biosynthetic gene cluster (*pkI*) from *Streptomyces venezuelae* ATCC 15429.
 (Updated on 15-SEP-2003 to standardise OS field)

CC biologically active variant, where the nucleic acid sequence is not

Db 9351 GGG 9353

Search completed: April 7, 2006, 03:16:47
Job time : 346.081 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 34751 Seconds
(without alignments)
12149.485 Million cell updates/sec

Title: US-10-611-442-1

Sequence: 1 ggcgccttcgcatcaccatcg.....gtcgagcaccggaagatcc 9024

Scoring table: IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	253.2	2.8	1521	10	CL440273 L001 PA_0
C 2	207.8	2.3	1135	9	BZ558756 PA98401.4
C 3	141.6	1.6	584	9	BZ298798 CG4587.F1
C 4	127.2	1.4	719	9	BZ562892 BZ562892 pacs2-164
C 5	118.4	1.3	1628	9	CG757066 P052-2-A0
C 6	116	1.3	987	9	CG142184 NDL.80J12
C 7	112	1.3	2332	10	AG363333 Mus muscu
C 8	110.4	1.2	847	8	DR887861 JGI CUNH1
C 9	110.4	1.2	2243	10	AG381986 Mus muscu
C 10	107.4	1.2	2332	10	AG363333 Mus muscu
C 11	106	1.2	1956	10	CG754548 P050-1-A1
C 12	102.2	1.1	1462	10	AG41877 Mus muscu
C 13	101.4	1.1	1552	10	AG430101 Mus muscu
C 14	101.4	1.1	1628	10	CG757066 P052-2-A0
C 15	100.2	1.1	1131	10	AG042920 Pan trogl
C 16	100	1.1	1569	9	AG341503 Mus muscu
C 17	99.4	1.1	874	9	BZ56165 pacs2-164
C 18	98.6	1.1	853	9	CG866089 NDL.125H8
C 19	96.4	1.1	1421	7	CK209412 FGA502117
C 20	96.4	1.1	1569	10	BZ569488 pacs2-164
C 21	96.4	1.1	1569	10	AG341503 Mus muscu
C 22	96.2	1.1	1853	10	AG382854 Mus muscu

C 23	95.8	1.1	1450	10	AG429738 Mus muscu
C 24	95.8	1.1	1821	10	CL090560 ISB1-17N1
C 25	95.6	1.1	1474	10	CL500927 SATL_68_H
C 26	95.4	1.1	1552	10	AG430101 Mus muscu
C 27	95.4	1.1	1598	10	AG030579 Pan trogl
C 28	95.4	1.1	2243	10	AG381986 Mus muscu
C 29	94	1.1	1406	9	CG756569 P051-4-B0
C 30	93.8	1.0	975	9	BZ675691 P051-4-B0
C 31	93	1.0	1285	9	BZ567880 pacs2-164
C 32	92.8	1.0	1885	10	AG363563 Mus muscu
C 33	91.8	1.0	1189	10	AG032118 Pan trogl
C 34	91.4	1.0	846	10	CL422977 L001 PA_0
C 35	90.6	1.0	1610	9	BZ569386 pacs2-164
C 36	90.4	1.0	1462	10	AG41877 Mus muscu
C 37	90.4	1.0	1780	10	AG448243 Mus muscu
C 38	89.8	1.0	1126	10	AG064051 Pan trogl
C 39	89.6	1.0	1069	8	DR739591 FGA508480
C 40	89.6	1.0	1542	10	AG032943 Pan trogl
C 41	89.6	1.0	1542	10	AG032943 Pan trogl
C 42	89.6	1.0	1821	10	CL090560 ISB1-17N1
C 43	89.4	1.0	2354	10	AG060256 Pan trogl
C 44	89.2	1.0	814	2	BE055163 GA_Ea003
C 45	89	1.0	1151	7	CK210749 FGA502257

ALIGNMENTS

RESULT 1
LOCUS CL440273/c
DEFINITION L001 PA_0005_L16 pooled genomic library of 12 strains of Pseudomonas aeruginosa
L001 PA_0005_L16, genomic survey sequence.
ACCESSION CL440273
VERSION CL440273.1 GI:45577320
KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1521)

AUTHORS Shen, K., Sayeed, S., Hu, F.Z., Antalis, P., Gladitz, J., Ahmed, A., Hayes, J., Johnson, S., Dopic, R., Preston, R.A., Erdos, G., Post, J.C. and Ehrlich, G.D.

TITLE Sequence from pooled genomic library of 12 isolates of Pseudomonas aeruginosa

JOURNAL Unpublished (2004)
COMMENT Other GSS: L001_PA_0026_A07

CONTACT: Shen K
Center for Genomic Sciences
Allegheny-Singer Research Institute
320 East North Avenue, Pittsburgh, PA 15212, USA
Tel: 412 359 5097
Fax: 412 359 6995
Email: kshen@panh.org
Complete insert sequence
Seq primer: M13 Reverse, and T7
Class: shotgun
High quality sequence stop: 1521.

FEATURES

source
1..1521
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="pooled"
/db_xref="taxon:287"
/clone="L001 PA_0005_L16"
/cell_type="bacteria"
/clone_lib="pooled genomic library of 12 strains of Pseudomonas aeruginosa"
/note="Vector: pCR4-TOP0, using TOPO Shocgun method; 12 strains include: 1) 11 isolates from clinical samples of pediatric otitis media with effusion: Pitt 'A', Pitt 'B',

pltc 'C', pltc 'D', pltc 'B', pltc 'F', M18851, M18858,
W27912, W27931, and W28869; 2) 1 strain from American
Type Culture Collection (ATCC): 27853 "

[illegible][illegible]

Db 576 GCAAGCCGACCTCGTTGAGCTTGCCGTTGATGCGGGCATGACACCGTACCTGCTCG 517
Qy 4410 CCGATGCCGAAGTTGTGTCATGCGCCGGATCGGTTGCGGCAGAGCGCGTGTGTAACC 4469
Db 516 ACBAAGCCCAAACTTTAAGGTGTGAGCGTTGCGGGGTGTTTGTGTCAGACAGATG 457
Qy 4470 ACCGCCCGCCCTCGAAGGCGGTGACCGCTTGCGGTGCGTGAAGTGAACACTCGGCG 4529
Db 456 ATGCAACCGCCCTCGAAGGTGTGGAACACTTGGTGCGCATGGAAGTGAAGTGTGAGG 397
Qy 4530 TCACCGGATCCGCCACCGGACGTCACCGGTGTGACGCCAGGCGGTGCGGCGCTCG 4589
Db 396 TCCCGGTGCGCAAGCAGCTGCGCGCCGCTCGCGCAAGGAGCATCGCGGCAATCG 337
Qy 4590 AAGAAGCTTGAAGCTGTGTGCGGGCGGATCTTGCCGACCGCTTCCAGAGTGTGCT 4649
Db 336 TGAATGACCTTCAGGTGTGTGAATATTCGGGATCCGCGATCCCTCGAATCGCAGGG 277
Qy 4650 CCGCCCCACAGGTGTATCGCCGACGATCGCCGCTGTCGCGGGTGAACAGCGCGCGACG 4709
Db 276 TTGCGGTAGCAATGACCGCGCAGAGATGCGGTGTCTGCGGGGTGAATCGCGCTGATC 217
Qy 4710 TGTCTCGGGTGCACGACGCGGTGCGCGGTGCACTGCGAAGAACACCGGTGTGATGCTG 4769
Db 216 CCGGACGGGTGAGATTCATGAGTCTGCTGATGTCACGAGATCGGCTTGAAGCCG 157
Qy 4770 AGCCACTGCGCGCGGTGCGCGGTGCGCGAGTCAATGACCGCATGATCACTTCAACCG 4829
Db 156 TTTCACAGCAGCGCGGTGCTGTGTGCGACGAAAGATGACGCTGTGATCACTCGCG 97
Qy 4830 GTGACGTCAACCGCGCCGACGACAGTTCAGGGCGACGCGTGGCGTTG 4877
Db 96 GTACCCCGCAATGCTGCAAGCGCGGTGATGAGCGCCAGGGTGGCGTTG 49

RESULT 3
BZ298798/c 584 bp DNA linear GSS 31-OCT-2002
LOCUS CG4587.r1 Candida glabrata Random Genomic Library Candida glabrata
DEFINITION genomic clone CG4587, genomic survey sequence.
ACCESSION BZ298798
VERSION BZ298798.1 GI:24442259
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Bukhryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
AUTHORS 1 (bases 1 to 584)
TITLE the 'asexual' pathogenic yeasts for a complete sexual cycle in
JOURNAL Genome Biol. 4 (2), R10 (2003)
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.
FEATURES
source 1..584
Location/Qualifiers
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG4587"
/clone_1ib="Candida glabrata Random Genomic Library"

ORIGIN
Query Match 1..6%; Score 141.6; DB 9; Length 584;
Best Local Similarity 56.2%; Pred. No. 5.2e-15;

Matches 289; Conservative 0; Mismatches 219; Indels 6; Gaps 1;
Qy 4411 CGATGCCGAAGTTGTGTCATGCGCCGGATCGGTTGCGGCAGACGCGGTGTGTAACA 4470
Db 557 CCAATGCCCTGTGTGCGACAGCGGACCGATGTGCGGGCGAACCTTGCTGTGTCGCAAGACA 498
Qy 4471 CCGCCCGCCCTCGAAGGCGGTGACCGCTTGCGGTGCGTGAAGTGAACACTCGGCGT 4530
Db 497 CCAATGCCCTGTGCGGACGATATTTCTTCTCGGATGAGGCTGAACATGCGCAACG 438
Qy 4531 CACCGGATCCGCCACCGACGTCACCGCTGTGACGCCAGGCGGTGCGCGCTGCA 4590
Db 437 TGCAGAACAGACCGGACCGGACGCGCTGCAAGGCTGCGCGCTGCGCGCGCTCT 378
Qy 4591 AAGAAGCTTGAAGCTGTGTGCGCGCGGATCTTCCGACGCGCTTCCAGAGTGTGCT 4650
Db 377 CGAAGGACAGACACGCTCTCTTCGCGCATCTCTCGAGCGCGCTCATGTCCGCGGT 318
Qy 4651 GGGCCCAAGGTGTATCGCCGACGATGCGCGGCTGTCGCGGGGTGAACAGCGCGGACGT 4710
Db 317 GACCTGAGAGGTGCAACCGCATGATGCTTCTGTGCGCTGCGATGCGCGCGACCG 258
Qy 4711 GGTCCGCGGTGACACGACCGGTGCGCGGTGCACTGCGAAGAACACCGGTGTGATCCGA 4770
Db 257 CGTCCGCGACAGGACGAGAGTGTGCGCTCGATGTCCGGAACACCGGGTGCAGCG 198
Qy 4771 GCCAGTCCCGCGGTGCGCGGTGCGCGGAGTCAATGACCGCATGATCACTTCAACCG 4830
Db 197 TCAGCGGACGAGTGTGCGGTGCGCGGACGTAACGACGAGATCACTTGCCTGCG 138
Qy 4831 T-----GACGTCAACCGCGCGCGGACGAGTTCAGGGCGAACGCGGTGCGCAAGTGG 4884
Db 137 CCGGCGCATGTCCTCGCGGAGAGACCGAGGTGAGACCGAGAGTCCCGAATTGACCG 78
Qy 4885 CGATGCAATGCCGTACCCCGACCAAGTTCGCGAC 4918
Db 77 CGACGCAAGTGGCGCGCGCGGACGAGCAGTTCGCGC 44

RESULT 4
BZ562892 719 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3982.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3982, genomic survey sequence.
ACCESSION BZ562892
VERSION BZ562892.1 GI:27185631
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.B., Sims,E.E., Hastings,M.,
AUTHORS 1 (bases 1 to 719)
TITLE Pseudomonas aeruginosa library
JOURNAL Whole-genome-sequence variation among multiple isolates of
Pseudomonas aeruginosa library
COMMENT U. Bacteriol. (2002) In press
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..719
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_3982"
/clone_1ib="pacs2-164"

```

/notes="clinical isolate 2-164 Whole genomic shotgun
library."

```

Query Match	Score	DB	Length
1.4%	127.2	9	719

Best Local Similarity 56.04; Pred. NO. 2e-12;
Matches 262; Conservative 0; Mismatches 203; Indels 3; Gaps 1,

QY	4410	CCGATGCCGAAGTTGTGATGAGCCCGGATCCGTTCCGACACAGGCCGTGTCGATGAC	4469
Db	168	CCGAGCCCGTGAATTGAGCGCGCGAGACCCTTCCGCGAGAGGACTGTCTCCCTGACG	227
QY	4470	ACCGCCCCGCGCTCGAAGACGAGTACCGCGCTTGCTGCGTGAAGCTGAACACTTCGCG	4529
Db	228	AACATCGCGCTTCGGCGACAGTCAAGTTCTTGATCGGTGGAAGAAAATGACGCGTG	287
QY	4530	TCACCCGATCCGCCACCGGACATCCCGTCGTGACGCCAAGGCGCTTGCGCGCGTGC	4589
Db	288	CCGCGGGA -- GCTTACCGCGCGCGCCCGCGTATTCGTCTGCGGACCGCTTGACCGCGTCC	344
QY	4590	AAGAAGACTTGAACCTGTGTGTCGCGCGCGCATCTTCGACAGCGCTCCACAGTGTG	4649
Db	345	TCGACCAAGGCGAATCCGCTGCGCGCCCGCGACCGTGCAGCAGGCTTCAGATTCAGAGTG	404
QY	4650	CGGCCCAAGGTGTACGCGCGACGATCGCGCGCGCTTTCGCGGAGTGAAGACCGCGGACG	4709
Db	405	GAGCCCGGTAGTGAACCGGACGATGCGCGCGTGCACCGGCGCATGCGCGCTTCCACT	464
QY	4710	TGATCCGAGTGCACAGACCGGATGCGCGGATGCACTGCGCAAAACACCGATGTGATGTCG	4769
Db	465	GCGTCGCCGAACACATCAGGAGTGTCCGCGTGCAGCTGACGAAACACCCCGGATGCGCA	524
QY	4770	AACCACTCGCCGCTGCGCGGTGCGCGGAAGTCAATTGAACGCGATATCACTTCACCG	4829
Db	525	AGCAGCGTATGACGTTGTCAGTGTGACCCAAATTAGGAGCGGCTTGATCACTTTGTTA	584
QY	4830	GTGACGTACCGGCCCCGACGACCAAGTTCAAGGCGAAGCGTGGCGGTG	4877
Db	585	GCAGAGCCATATCCCTTAGGCCACGTGCGTGAATACATATGAGCGCGGTAG	632

[illegible]

```

/dbc xref="taxon:54126"
/clonelib="Ppa EcORI BAC library"
/notes="The library was generated by a partial digest of
the genomic DNA with EcORI and cloning into the BAC
vector."

```

Query Match	Score	DB	Length
1.3%	118.4	10	1628

Best Local Similarity 44.1%; P-vec. NO. 1,35-11;
Matches 639; Conservative 0; Mismatches 797; Indels 14; Gaps 5;

QY	7543	CACGAGACCGGCGCGGAGAGAGTCTGAGCAGATCCGAGATCTACGACAGGACGACGAGATC	7607
Db	1612	CGCGGCG	1553
QY	7603	GCCCCGAGAGATGAGAGATGAGCGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	7662
Db	1552	NCCCCGCCCCG	1493
QY	7663	TTCTTACGCGCCGACACGAGAGACGCGCGCGCTGAGTAC-----AGTGAACGCGGACTGCTCC	7717
Db	1492	CCCCCG	1433
QY	7718	AGCAGCGGTGCGCGCAGGCTGTGTGAACGAGCTGCGGAGTGTGCGCTTGTGATCTGTGAACGACT	7777
Db	1432	CG	1373
QY	7778	TGCGCCGCGGAGAGTACCCCGGTGCTGCGCGCTGCGGTACCGCGCGCGCGACACTGAGGAGCGTGAAC	7837
Db	1372	GCG	1313
QY	7838	CGACCGGTCTCGCGGTCTTGAGACCTTGCGCGACCGCGGATATGCTTGAGCGCGCAGGTACGCC	7897
Db	1312	CG	1253
QY	7898	CGCAACAGCTGCGCGGTGAACGAGCGCGCTGACCGCGCTTGACAGATGAGCGCGGTCA	7957
Db	1252	GCCCCGCGCGGAGAGCG	1193
QY	7958	CGCGCGCGGTGCGGAGACGCGCGCGGTGTGCTGTGAGGAGAGTGTGAGAGTCTGAGCGCGCAACGCG	8017
Db	1192	CGGAGGAGCG	1133
QY	8018	TGAGGAAACGCGCGTCTGTGCGCGGTCAACGAGCTTCCGGAATTGAGCGGCAACGACTTGCGACG	8077
Db	1132	GCG	1073
QY	8078	ACCGGAGAAACCGCGAACCCGAGTGTGTGACAGAGATGTGTGCGGAGCGAAGTCCGCGGCTCAAC	8137
Db	1072	CCCCCCCCCG	1013
QY	8138	TGGAACGCGCGACGCGCGCGCGGTGAGACGCGCGAGTGTGAGCGCGGAGTGTGACGTCCCGAACCGGTG	8197
Db	1012	CG	953
QY	8198	GCGAGGTGACAGTGTGTGTGTGCGCGCGCGGAAACGCTGATCCGAGAGTCTTACACGATCCCG	8257
Db	952	CCCCCGCGCGCGCGCG -CCCCCG	894
QY	8258	ACCGGTTGAGCGTGAACCTGTGAGCGCGCGCGCGCGCGCGAGATCTGTGTGTCCCGCGCGGAGCTCG-	8316
Db	893	CG	834
QY	8317	-CGCGCGCGCGACTGACGCGCGCTGT	8375
Db	833	CG	774
QY	8376	CGCGCGCGGT	8435
Db	773	CG	714
QY	8436	CGTGCCTCGTGTCTTACCGCGTGTGCGCGCGGTGTGAGCTGTGAGGAGAAAGATGTGCGGTGCG	8495

[illegible]

RESULT 6	CC142184	987 bp	DNA	linear	GSS 16-APR-2003
LOCUS	CC142184				
DEFINITION	NDL.80J12.T7 Notre Dame Liverpool Aedes aegypti genomic clone				
ACCESSION	CC142184				
VERSION	CC142184.1	GI:30011239			
KEYWORDS	GSS.				
SOURCE	Aedes aegypti (yellow fever mosquito)				
ORGANISM	Aedes aegypti				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Culicinae; Culicini; Aedes; Stegomyia.				
REFERENCE	1 (bases 1 to 987)				
AUTHORS	Loftus, B., Shetty, J., Knudson, D. and Severson, D.				
TITLE	BAC end sequencing of Aedes aegypti				
JOURNAL	Unpublished (2003)				
COMMENT	Other_GSSs: NDL.80J12.SP6				

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: entc@tigr.org
 Library was provided by David Severson
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1..587

FEATURES	location/Qualifiers
source	1..987
	/organism="Aedes aegypti"
	/mol_type="genomic DNA"
	/strain="liverpool"
	/db_xref="taxon:7159"

```

/c/clone="NDL.80012"
/c/clone.lib="Notre Dame Liverpool"
/note="Vector: PECBAC1 Site_1: Hind III, The library was
prepared from whole body tissue of newly hatched L1 larvae
by David Severson at the University of Notre Dame and
Hongbin Zhang"

```

Query Match	1.3%	Score 116;	DB 9;	Length 987;
Best Local Similarity	47.8%;	Pred. No. 2e-10;		
Matches 367;	Conservative 0;	Mismatches 400;	Indels 1;	Gaps 1;

QY	4135	CCGGGCCAGGATGCGCTGCAAGCCGGTCAAGCGATGGCGGTGCGTCCACCA	4194
Db	202	CCGGGCTTTGAGTCGTCAATACAGTTCAATCCCGCTGACCGGATAGGATCTTTCACCA	261
QY	4195	GGATGATCAAGTACTGGTAGTTGCTCTCTGTCGGGCGGGAGCGAGTGCACGGTGAAGC	4254
Db	262	GAATAGGAAATAGGCAATAGTTTGCAACTTTTCCCGGCAATGACAGACACTGATATTC	321
QY	4255	CGGTAGTGGCGAGCTCCCTGCTGATAGACGGTGGATGACCCCGTTGTGCTCCG	4314
Db	322	CTGGACAACCTCCAGAAAGCTCCGGAATACAGTTCACTATATTTCTTCGCTTTAGCAAA	381
QY	4315	TCTGGGCGAAAGCGCTCAGAGGAGGTGAGCCCAATGGCCGCGGCACTCGCTCATCTTGC	4374
Db	382	TACATCATGATGCGTTTACGTGACGAAAGGCCAGCGGGCAATTTATCTCACTCATTTTCC	441
QY	4375	CGTTGATCCGATCTCGGTGACCACTTGTCCGGGCGATGCGCAGATTGTGATGAGCCC	4434
Db	442	CATGATTTCCGGGCTTACCAAGTGGTTTCAATTGACAAAGCCGAAGTTCTTAAAGAT	501
QY	4435	GGATCCGTTGGGCGACAGGCGGTGTCGGTGAACAACCGCCCGGCTCTCGAAGCGGTGA	4494
Db	502	CGAATCCCTCTTAACTTGTGGCAATCCCGGCGAATGATGAGTGTCCGCTTCGAAATGTTTGA	561
QY	4495	CCGCTTGTGTGGCGGTGAGCTGAAACACTCCGGCGTACCGGAATCCGCCACCGGAAGTCC	4554
Db	562	ATACCTTGTGTGATGAGTGAAGCTCAACCCGATTAATACATAGCGCAAAATACCTTCCC	621
QY	4555	CACCCGTCGTGACAGCCAGAGGGCGTGGCGGCTGGAAGAAAGCTTGACCTGTGTGTGCG	4614
Db	622	CATCATCAAGCAACCGAAAGCATGGCGGCAATCATGATCAACCGTAAGTTGATAT	681
QY	4615	CGGCGATCTTCGCAAGGCGCTCCACAGGTGCTGTGCGGCCACAGGTGATCGCCAGCA	4674
Db	682	CTGGAATCTTTTGTATCTTATCTACATCGCAAGGATGTCATTAACAATGACAGGCAATGA	741
QY	4675	TTCGCGCCGGTCTCGGGGGTGAACAGAGCGGCGGACGTGTGTCGGGTGCAACAGACCGGTGC	4734
Db	742	TGGCTGTGCTGATAGGGGTATGATGCTGCTTATCTGTCTCGGATTCATGTCAGCGTTC	801
QY	4735	CCGGGTGCAAGTGCAGAAACACCGGTGTGATGCTCGAGCCAGCTGCGCGCGTGC	4794
Db	802	GCGGATGATGATGAGCAACCGGCTTATGATGCAATTCACAGCAATGAATTCGCGCATGG	861
QY	4795	CCGGAAAGTCAATTGACGCGATATCACTTACCGGTGACGTCAACGGCCCGACAGCA	4854
Db	862	-CAACAATGAATAGGGGTGCTGTTATGACTTCAACCGGTTATTTCAATGCTTGAAGTGG	920
QY	4855	GTTTCAAGGGGAGCGTGTGCGTTGACAGTGGCGATGCAAGTGCCTTACCC	4902
Db	921	TCAACAAGCGCAATGTTTCAATTGAAAAACAAGCCAGGTGATGTACAC	968

RESULT 7	AG363333/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
	AG363333	2332 bp	DNA	linear	GSS 21-DEC-2004	
	Mus musculus	molossinus	DNA,	clone:MSWg01-165A22.TJ,	genomic survey	
	sequence.					
	AG363333					
	AG363333.1				GI:47974538	
	GSS.					

[illegible]

Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel.: 925 296 5600
 Fax: 925 296 5710
 Email: cdaa@jgi-psf.org
 Tissue Procurement: Dan Buchholz (Yun-Bo Shi Laboratory, NIH)
 cDNA Library Preparation: DOE Joint Genome Institute
 http://www.jgi.doe.gov
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
 Clone Distribution: I.M.A.G.E. Consortium/LLNL:
 http://image.llnl.gov
 Naming Conventions: EST name is generated by the concatenation of
 the UCI Clone id and the direction of sequencing. The suffix ',rev'
 indicates a reverse sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Plate: CUNH 0013 row: 9 column: 23
 High quality sequence stop: 815.
 Location/Qualifiers
 1..847

```

/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGR:7762993"
/tissue_type="whole embryo"
/dev_stage="Metamorphic (st.62)"
/lab_host="Electromax DH10B"
/clone_1lb="NH XGC tropMet2"
/notes="vector: pSPORT1; site_1: SalI; Site_2: NotI. This
library was made from dr primed cDNA and cloned into
Invitrogen pCMVSPORT6 vector. The work was done at DOE
Joint Genome Institute. Poly A RNA were primed with 5'
GACTAGTCTTAGATCGGAG CGGCCCGCCCTTTTCTTTT 3'. cDNA
were ligated to SalI adapter (5' TCGACCAAGCGGTGG' and
5'CGGACGCGTGG)', digested with NotI, size fractionated in
1.1% agarose gel electrophoresis and ligated into NotI and
SalI digested pSPORT1 vector."

```

Query Match	1.2%	Score 110.4	DB 8	Length 847
Best Local Similarity	50.3%	Pred. No. 2.1e-09		
Matches 307	Conservative 0	Mismatches 231	Indels 12	Gaps 1
QY	4297	CCCGGTTGTCCTCCCTGATCTCCGCGCAACGGTCGAGGGAGGTGAGGCCCACTGACCGCGG	4356	
Db	613	CTGGCGCATGAGCGGTATCTTCTCGATATATGGGCAATTTGGCCCAAGCCGATGGCCGCTG	554	
QY	4357	CGCACTGCTCATTTTSCCGTTGTCGCAATCTCGGTGACCACTTGTCCGGGCCGATGC	4411	
Db	553	CGATGTGCGTCATGTTGAACTTGGCCACCGAGTACATCCAGTCCAGGCCGTCGAAGCCGG	494	
QY	4417	CGAAGTTGTGATGAGCCCGGATCCGCTTCGGCCAGCAAGC-----CGTCGTGG	4466	
Db	493	TACGGGTAAACGCTTGAGCGCATATTTTTCGCGACAGCGCGCTTCTTGACGTTGTTC	434	
QY	4465	TGACCAACGGCCCGCCTCGAAGCGGTGACCGGCTTGTGTGCGCTGGAAGCTGAACCT	4524	
Db	433	ATACACAGCGACCGCCCTTCGGATGACGTGATGTTCTTGTTCGCTCGAAGCTGAACGCA	374	
QY	4525	CGGCGTCACCGAATCCGCCCAACCGGACGTCCACCGTCGTGACGCCCAAGGCGTGGCGG	4584	
Db	373	CGAATGCGCGGTAGACCAATGCGTCAACCATCCAGCTGGAAACCAAGGCGCTGGCGG	314	
QY	4585	CGTCAAGAGAGGTTGACCGGTGTGTGCGGCGGATCTTTCGCAAGCGCTCCACAAGTG	4644	
Db	313	CATTTCAAGAGTGCAGGTTGTATCTTTGTGGCAACGCGGTAGAGCAAAAGCATGTCCA	254	
QY	4645	CTGGTCGGACCCCAAGGTGTACGCCCGACGATTCGCGCGGTCTGGGGGAGTGCAGACCGG	4704	
Db	253	GCGGCAGACCGGCGCAGGTATACCGGAGTATTTGCTTGTGTGCGGGGTGATCGCGCTT	194	
QY	4705	CGACGTGTGTCGGGTCGACCAACCGGTGCGCGGGTTCACGTTCGACAAACCGGTTGA	4764	

Db 193 CCACCTGGCCCAAGTCGATATTGCGGGTCACCGGGTCGATATCGCGCAACCGGCGTGG 134
QY 4765 GTCCGAGCCAGCTCCGCGGTGGCGGGTGGCCCGGAAAGTCATTGAGCGCATGACTT 4824
Db 133 CGCGGACTTCCAGAAATACGTTGGCTAGCTACCCGAAATCGGGGTGTGATGACTT 74
QY 4825 CACCGGAGTCAGTCACCGGCGCGGACGACAGTTCCAGGCGAGCGGTGGCGAGTGG 4884
Db 73 CGTCCCAAGGCCCGCATCCGCGCATGGCGCAAGCGAATCTCATGTGTGCAAGTGGCGAGT 14
QY 4885 CGATGCAGTG 4894
Db 13 TGAAGTGCG 4

RESULT 9
AG381986 2243 bp DNA linear GSS 21-DEC-2004
LOCUS AG381986
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-192D21.TU, genomic survey
sequence.
AC381986
VERSION AG381986.1 GI:47993191
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS 1
Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T.,
Brawa, K., Saito, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and
Shiroishi, T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
Sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
AUTHORS 2 (bases 1 to 2243)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail: hattori@gsc.riken.jp, URL: http://bgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

PRIMERS
Sequencing : TU
LIBRARY : PBACe3.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source 1..2243
Location/Qualifiers
1..2243
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-192D21.TU"
/sex="male"
/issue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 1.2%; Score 110.4; DB 10; Length 2243;
Best Local Similarity 42.5%; Pred. No. 2e-09;
Matches 514; Conservative 0; Mismatches 693; Indels 2; Gaps 2;

QY 7490 CCGCGGAGACGGAGACCGTACGCCAGTCGAGCCGGAACGTGGGTGTGTGACACCGGA 7549
Db 974 CCCCCCCCCCGGGGCGCNCNCGCGCNCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1093
QY 7550 CCGCGCGCGAGGTGTCTCGGATCCGAGCTTCAACCGACCGGCGCGGACGCTGCGCGT 7609
Db 1034 GCG 1093
QY 7610 GATGACAGGTGACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7669
Db 1094 GGGGGCG 1153
QY 7670 CCGGACCGAGAGACGCGCGCGTCTCGGTGACAGTGAACGCGGACTGTTCACAGCGGTG 7729
Db 1154 CG 1213
QY 7730 CAGAGCTGTGACCGAGCTGAGGTGCGCTTCGATCTGTGAACGACTTGTGCGCGGAG 7789
Db 1214 GCG 1273
QY 7790 TCCCGGTGTGCGCGCTCGGTAACCGCGCGCGCACTCAAGGGGTGGAACCGGACCTGTC 7849
Db 1274 CGCCCG 1333
QY 7850 GATCTTGAACCTCGCGGACCGCGGTATGCTTGAACGCGGATCAAGCGCGCAAGCTCG 7909
Db 1334 CG 1393
QY 7910 CGGTGACCGAACAAGCGCTGACCGCTTCAAGATCAACGCGGTACCGCGCGTTCGG 7969
Db 1394 CG 1453
QY 7970 ACGCGCGGT 8029
Db 1454 GGGCGGACG 1513
QY 8030 TCTGTGCGCTGACCGGACTTCCGAACTGCGCGGACGACTTGGACGACCGCGGAGACG 8089
Db 1514 CG 1573
QY 8090 CGACCGGTGTGTGACGAGAGGTGTGCGGACAGTCCCGCGCTTCACTGGAACCGGCA 8149
Db 1574 GGGCGGGGCG 1633
QY 8150 CCGCGCGGTGCG 8208
Db 1634 CG 1693
QY 8209 GT 8268
Db 1694 GCGGCG 1753
QY 8269 GTGACCGGTGCGCGCGGACGCGCGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 8327
Db 1754 GGGCGGACG 1813
QY 8328 CTTGACGCGCTGT 8387
Db 1814 GGGGCG 1873
QY 8388 GCGCGCGGTGCGCGGTTCGCGGCGCGGTGATCGAGATCGATCGATCGATCGATCGATCG 8447
Db 1874 CNGCG 1933
QY 8448 TCTGACCGGTGCGCGGTGAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8507
Db 1934 GCG 1993
QY 8508 TGGGTGTCAACAGCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8567
Db 1994 CG 2053

QY 8568 GACACGAGTACGAGTCTGCTGCTCGCGGCGCTGACCGACGAGTCAACGAGTCCCGTTC 8627
DB 2054 NCSCNCGNCGNCGNCGCGCCCGCCCGGANNNGCGCGCGCCCGCGGCGGAGNCGCGC 2113
QY 8628 TAGCCCCCGTCCCGTCTGAGTGAAGTCTGAGTGTGACAGCCCGACGCGGCGC 8687
DB 2114 GCGCNCNAGGCGNCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2173
QY 8688 AGGACATCG 8696
DB 2174 CCGGCGCGC 2182

RESULT 10
AG363333
LOCUS Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey
DEFINITION AG363333 2332 bp DNA linear GSS 21-DEC-2004
ACCESSION AG363333
VERSION AG363333
KEYWORDS GI:47974538
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS 1
Abe,K., Noguchi,H., Tagawa,K., Yuzurika,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shirotani,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15374823
REFERENCE 2 (bases 1 to 2332)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunya Abe (abe@rc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY : DBACE3.6
Vector : EcorI
R.Site 1 : EcorI.
R.Site 2 : EcorI.
FEATURES
source Location/Qualifiers
1..2332
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-165A22.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC library"

ORIGIN
Query Match 1.2%; Score 107.4; DB 10; Length 2332;
Best Local Similarity 41.9%; Freq. No. 7.1e-09;
Matches 709; Conservative 0; Mismatches 948; Indels 35; Gaps 5;
QY 1194 CCTCCGACACCGCACCTTCTCTCCGACCCCGGTGATGAGGAGCGCGACCC 1253

DB 616 CCNCGNCGACCCACCTTCTCCGCGGCGCGCTGCGCGCGGAGACGCGCGCGCGC 675
QY 1254 GAGCGCGGAGTATCAAGAGTTCAGACCGCGCGGAGACCGGAGCCCTGCGAAGTCT 1313
DB 676 GCGTCCGCGTCCCGCGCAACCGCGGATGCGCGCGCGCGCGCGCGCGCGCGCG 735
QY 1314 CAGCAGCGCGCTTCAACCGCGGTACGATTCGCGGACCTCGAACCGCGC-----ATCGG 1365
DB 736 GCGGCG 795
QY 1366 GAGGTACCGCGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1425
DB 796 CCG 855
QY 1426 GCGTTCGCGTCTGCGGTACGATCTGCTGCGGAGTCTGCGGAGTCTGCGGAGTCA 1484
DB 856 ACCCG 915
QY 1485 CAGGAGTTCGAGTCACTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1544
DB 916 CCG 975
QY 1545 TCCGCGCGTGTGCGAAGCGATCAATGAGGTGCGAACCGCGTCACTTACCTGCTCGA 1604
DB 976 CCG 1035
QY 1605 CAGGTTCGCGAAGCG 1664
DB 1036 CCGCGCGCGTGTGCG 1095
QY 1665 CAGGTTCGCGAAGCG 1124
DB 1096 GCG 1155
QY 1725 GCTCGCGCGCGCACATCACACGACCGCTCTGCTGCGCGCAATGCTCGGACCTCGACGA 1784
DB 1156 GCG 1215
QY 1785 G-----CAACCGAGTACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1827
DB 1216 GCG 1275
QY 1828 ATCGCGCGATCATCGAGAGGTGTTCCGTTCCGCGCGCGCGCGCGCGCGCGCGCG 1887
DB 1276 CCG 1335
QY 1888 ACCACGACGAGGCGCGCACACCGTCTGCTGCGGAGTTCGAGATCCCGCGCGAGTCAAT 1947
DB 1336 GCG 1395
QY 1948 ACTGCGGTCTCTCGGCGAACCGCGATCTCTGCGCGATCCCGACCGGACGATTTCAG 2007
DB 1396 GCG 1455
QY 2008 CCGTCCCGAAGATCGGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2067
DB 1456 GCG 1515
QY 2068 CTCGCGGCG 2127
DB 1516 GCG 1575
QY 2128 TAGCGTCACTGCGCGTTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2187
DB 1576 GCG 1633
QY 2188 CTCGCGACCG 2247
DB 1634 CCG 1693
QY 2248 CCTTCGCGTCCGACCG 2307

QY	8256	CGACCGGTTTCACGCTGGAACCGTGGGCGGGAGACCGCAGATCCTGTCCTCCGGCCGGCTC	8135		
Db	1751	GGCGGGGCGGGCGGGCGGGGGGCGCTCCGCGCNGCGCGCCNCGCCCGGGGGGCGCGCG	1810		
QY	8316	GCCCCGACCGACCTCGACCGCTCTGATGGCCACCTGTGACACGCGCGCGCTGCGGGCGCG	83757		
Db	1811	CGAGCGCCCG	18707		
QY	8376	CGCGCGGATGTTGCCCCCGGCTGTCTCCGTTCCGGGCGGGTGATCAGACGACGTGGTCACC	8435		
Db	1871	CGCCCGCGAGNGGCGGCGCGCGCGCGCGGGGGGGGGCGCGCGGGGCGCGCCCGCGCGCN	1930		
QY	8436	CGTGGGCC	8443		
Db	1931	CGGCGCGC	1938		
RESULT 12					
LOCUS	AG441877	1462 bp	DNA linear GSS 22-DEC-2004		
DEFINITION	Mus musculus molossinus DNA, clone:MSMg01-323H09.TJ, genomic survey				
ACCESSION	AG441877				
VERSION	AG441877.1				
KEYWORDS	GI:48084940				
SOURCE	GSS.				
ORGANISM	Mus musculus molossinus (Japanese wild mouse)				
REFERENCE	Mus musculus molossinus				
AUTHORS	Ezawa,K., Satou,N., Hattori,M., Sakaki,Y., Moriaki,K. and Shiroishi,T.				
TITLE	Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end				
JOURNAL	Genome Res. 14 (12), 2439-2447 (2004)				
PUBMED	15574823				
AUTHORS	2 (bases 1 to 1462)				
TITLE	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.				
JOURNAL	Direct Submission				
PUBMED	Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical				
AUTHORS	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
TITLE	1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa, 230-0045, Japan				
JOURNAL	(E-mail:hattori@gsc.riken.jp, URL:http://bgp.gsc.riken.go.jp/,				
PUBMED	Tel:81-45-503-9111, Fax:81-45-503-9170)				
AUTHORS	Clones are derived from the mouse BAC library MSMg01. For BAC				
TITLE	library availability, please contact Kuniyasu Abe (abe@crc.riken.jp).				
JOURNAL	The Institute of Physical and Chemical Research (RIKEN) 3-1-1				
PUBMED	Koyadai, Tsukuba, 305-0074 Japan				
AUTHORS	phone: 81-298-36-9189, fax: 81-298-36-9199				
TITLE	e-mail: abe@crc.riken.jp				
JOURNAL	PRIMERS				
PUBMED	Sequencing : TJ				
AUTHORS	LIBRARY				
TITLE	Vector : PBACe3.6				
JOURNAL	R Site 1 : EcoRI				
PUBMED	R Site 2 : EcoRI.				
AUTHORS	Location/Qualifiers				
TITLE	1. 1462				
JOURNAL	/organism="Mus musculus molossinus"				
PUBMED	/mol_type="genomic DNA"				
AUTHORS	/sub_species="molossinus"				
TITLE	/db_xref="taxon:57486"				
JOURNAL	/clone="MSMg01-323H09.TJ"				
PUBMED	/sex="male"				
AUTHORS	/tissue type="mixture of kidney and spleen"				
TITLE	/clone_lib="MSMg01 Mouse Male BAC Library"				
JOURNAL	ORIGIN				
PUBMED	Query Match				
AUTHORS	1.1%; Score 102.2; DB 10; Length 1462;				

[illegible]

DEFINITION P052-2-A02.za Ppa EcoRI BAC Library *Pristionchus pacificus* genomic,
genomic survey sequence.
ACCESSION CG757066
VERSION CG757066.1 GI:37985257
KEYWORDS *Pristionchus pacificus*
SOURCE *Pristionchus pacificus*
ORGANISM *Bukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus*
1 (bases 1 to 1628)
Srinivasan,J., Slinz,M., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Bumferr,J., van der Meulen,M. and Sommer,R.U.
An integrated physical and genetic map of the nematode *Pristionchus*
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RU
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Classes: BAC ends.
Location/Qualifiers
1..1628
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN

Query Match 1.1%; Score 101.4; DB 10; Length 1628;
Best Local Similarity 44.7%; Pred. No. 8.5e-08;
Matches 556; Conservative 0; Mismatches 669; Indels 20; Gaps 5;

QY 7448 GACGAGACGACGACCCCTTGATGATGAAAGCTTTCGAGCCGCGACCGAGACGCGG 7507
DB 310 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 369
QY 7508 AGCCAGTCCGCGCGCGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7567
DB 370 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429
QY 7568 CCGATCCGAGCTTTCACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7627
DB 430 CCGCGGCG 489
QY 7628 GCGCGCGCGCGCTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7687
DB 490 CCG 549
QY 7688 CGTCCGATGACATGACG 7747
DB 550 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 609
QY 7748 TGGAGGTCGCGCTTCGATCTCGTGAAGCACTTCCGCGCGCGCGCGCGCGCGCGCGCG 7807
DB 610 CCG 669
QY 7808 GTACCG 7867
DB 670 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 729
QY 7868 CCGCGGATGCTGCGACG--CCAGAGTCAAGCCCGCAACGCTCGCGCGTACCGAACAGCG 7925
DB 730 GCGAGNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 789
QY 7926 GCTGACCGCGCTCGACGAGATGACGCGGCTCAACCGCGGTCGAGACGCGCGGTCGTGCT 7985

DB 790 GCG 849
QY 7986 GGGAGTGTGGCGGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8045
DB 850 GCG 902
QY 8046 GCTTCCCGAATGCG 8105
DB 903 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 962
QY 8106 GAGAGTGTGGCGGAGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8165
DB 963 CCG 1022
QY 8166 CCGAGTGTGGCGGAGCG 8225
DB 1023 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1082
QY 8226 GAACTGATCCCGAGGCTTTCACCGATCCCGAGCGGTTTCAGCTGGACCGTGGCGCGA 8285
DB 1083 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1142
QY 8286 CCGCGAGATCTGTGCTCCCG 8345
DB 1143 GCGCG---CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1196
QY 8346 CACCTTGGCGACGTT 8405
DB 1197 CCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1256
QY 8406 CCGCGCGGATGATCAACGAGCTGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 8465
DB 1257 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1311
QY 8466 CGAGCTGTAG 8525
DB 1312 GCG 1371
QY 8526 TGTTCGAGGCTGTGCTCCCGCTCGAGAGCGCGCTTTCAGGCGCGCGCGCGCGAGAGAG 8585
DB 1372 CCG 1431
QY 8586 TCGCTTCGT 8645
DB 1432 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1491
QY 8646 GTGACGAGCGTGAACCTTGTGAGTGGCAAGCGCCAGCGCGGAGCGAG 8690
DB 1492 GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536

RESULT 15
AG042920/c 1131 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-021806.R, genomic survey sequence.
DEFINITION AG042920
ACCESSION AG042920
VERSION AG042920.1 GI:16571645
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes (chimpanzee)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Pan.
REFERENCE 1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1131)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbese@sec.riken.go.jp, URL:http://hsp.sec.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

COMMENT

PRIMERS

Sequencing: M13Rev
LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .1131

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-021B06.R"

/sex="male"

/cell_type="Lymphoblast"

ORIGIN

Query Match 1.1%; Score 100.2; DB 10; Length 1131;

Best Local Similarity 45.0%; Pred. No. 1.4e-07;

Matches 406; Conservative 0; Mismatches 488; Indels 9; Gaps 4;

```

QY 7498 CCGGAGACCGTACCGCAGTCGGGACCGGACGTCGGTGTGTCAGTCCAGCCGCGCC 7557
DB 1104 CCGGTCGTGTGTCGGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1045
QY 7558 GAGGTGTCCTGCGGATCCGGGCTTTCACCAACGACCGCCCGGACGTCCTCGGTGATCAG 7617
DB 1044 CCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986
QY 7618 GTGGCCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7677
DB 985 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 926
QY 7678 GAGGACCGCGCGCGTCGTGTGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7737
DB 925 GAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 867
QY 7738 GTGACCGAGCTGGGCTCGCGCTTGTGATCTGTGA-----ACGACTTGGCCCGGAGTCC 7792
DB 866 CGGCCNNGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
QY 7793 CGGTGTGTGGGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7852
DB 806 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
QY 7853 CTTGACCTTGGGCGGACCGCGGTTATGCTTGAAGCGCCAGGTCAAGCCCGCAACAGCTCGCGG 7912
DB 746 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
QY 7913 TGAACGAAACAGGCGCTGACCGCGCTTCAAGAGATGACGCGGTCAACCGCGGTGAGACG 7972
DB 686 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
QY 7973 CCGCGGTGTGTGTGGGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8032
DB 626 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
QY 8033 TGGCGGTGACCGAGCTTCCGGAATGCGGCGGACGACTTGGCGGACCGCGGACCGCGGA 8092
DB 566 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 509
QY 8093 CCGGTGTGTGTGACCGGAGTGTGCGGAGAGTCCCGGCGTCCACTTGGAAGCGCGCACCG 8152
DB 508 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 449

```

```

QY 8153 CCGCGTCCGACCGCGCGGTGGGCGGGGTGACGTCCTCCGACCGGTGCGAGGTGACAGTGG 8212
DB 448 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 389
QY 8213 TCGTCCCGCGCGGAAACCGTGAATCCCGAGATCTTCAACCGATCCCGAACCGGTTGAGGTGG 8272
DB 388 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 329
QY 8273 ACCGTGCGCGGACCGCGCGGAGATCTGTGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 8332
DB 328 CCGCGCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 269
QY 8333 ACGCCCTGTGTGACCACTTGACCGACCGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 8392
DB 268 GACGCGAGATATCTGTGTTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 209
QY 8393 GGC 8395
DB 208 CGC 206

```

Search completed: April 7, 2006, 14:04:12
Job time : 34756 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd

Run on: April 7, 2006, 02:41:57 ; Search time 3401.3 Seconds
(without alignments)
10613.880 Million cell updates/sec

Sequence: 1 gcgcgccttcgatcaaccatgg.....gttgagagcaaccgaggagatcc 9024

Sequence: 1 gcgcgccttcgatcaaccatgg.....gttgagagcaaccgggagagatcc 9024

Searched: 9267905 seqs, 2000278028 residues

Total number of hits satisfying chosen parameters: 18535810

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA New: *

- 1: /SDSS/p/odata/2/pubnba/US08_NEW_PUB.seq.*
- 2: /SDSS/p/odata/2/pubnba/US06_NEW_PUB.seq.*
- 3: /SDSS/p/odata/2/pubnba/US07_NEW_PUB.seq.*
- 4: /SDSS/p/odata/2/pubnba/PCT_NEW_PUB.seq.*
- 5: /SDSS/p/odata/2/pubnba/US09_NEW_PUB.seq.*
- 6: /SDSS/p/odata/2/pubnba/US09_NEW_PUB.seq.1*
- 7: /SDSS/p/odata/2/pubnba/US10_NEW_PUB.seq.1*
- 8: /SDSS/p/odata/2/pubnba/US10_NEW_PUB.seq.1*
- 9: /SDSS/p/odata/2/pubnba/US10_NEW_PUB.seq.2*
- 10: /SDSS/p/odata/2/pubnba/US10_NEW_PUB.seq.3*
- 11: /SDSS/p/odata/2/pubnba/US11_NEW_PUB.seq.2*
- 12: /SDSS/p/odata/2/pubnba/US11_NEW_PUB.seq.2*
- 13: /SDSS/p/odata/2/pubnba/US11_NEW_PUB.seq.3*
- 14: /SDSS/p/odata/2/pubnba/US11_NEW_PUB.seq.4*
- 15: /SDSS/p/odata/2/pubnba/US06_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	128.2	1.4	1694869	7	US-10-506-454-1690	Sequence 1690, App1
2	122.8	1.4	116856	14	US-11-143-960-1	Sequence 1, App1
3	121.6	1.4	1421	14	US-11-075-185-53	Sequence 53, App1
4	121.6	1.4	14122	14	US-11-075-185-2	Sequence 2, App1
5	118	1.3	767	8	US-10-497-135-17	Sequence 17, App1
6	115.6	1.3	1215	14	US-11-143-401-74	Sequence 74, App1
7	112.2	1.3	761	8	US-10-497-135-18	Sequence 18, App1
8	107.8	1.2	1236	14	US-11-143-401-127	Sequence 127, App1
9	92.6	1.0	1884	11	US-11-096-5684-20202	Sequence 20202, App1
10	92.6	1.0	88421	14	US-11-205-109-1	Sequence 1, App1
11	89.6	1.0	1412	14	US-11-075-185-2	Sequence 2, App1
12	89.2	1.0	2736	8	US-10-658-730-38	Sequence 38, App1
13	87	1.0	88421	14	US-11-205-109-1	Sequence 1, App1
14	86.6	1.0	1314	14	US-11-175-689-5	Sequence 5, App1
15	85.4	0.9	3240	14	US-11-052-5544-529	Sequence 529, App1
16	83.6	0.9	2724	8	US-10-517-939-359	Sequence 359, App1
17	81.2	0.9	1485	14	US-11-143-960-20	Sequence 20, App1
18	81.2	0.9	116856	14	US-11-143-960-1	Sequence 1, App1

C 19	80.4	0.9	6408	8	US-10-858-730-40	Sequence 40, Appl
C 20	79.4	0.9	3408	8	US-10-497-135-02	Sequence 2, Appl
C 21	79.4	0.9	41465	14	US-11-052-5544-532	Sequence 522, Appl
C 22	79.2	0.9	4983	14	US-11-052-5544-531	Sequence 521, Appl
C 23	77.6	0.9	5706	14	US-11-052-5544-519	Sequence 519, Appl
C 24	77.4	0.9	1593	8	US-10-858-730-138	Sequence 138, Appl
C 25	76.6	0.9	8651	8	US-10-432-483-48	Sequence 48, Appl
C 26	76.4	0.9	5706	8	US-10-497-135-1	Sequence 1, Appl
C 27	75.6	0.8	5706	8	US-11-052-5544-519	Sequence 519, Appl
C 28	75.4	0.8	4617	14	US-11-052-5544-530	Sequence 530, Appl
C 29	75.2	0.8	2736	8	US-10-858-730-38	Sequence 38, Appl
C 30	72.4	0.8	5679	14	US-11-075-185-36	Sequence 36, Appl
C 31	72.4	0.8	78869	14	US-11-075-185-1	Sequence 1, Appl
C 32	72	0.8	41465	14	US-11-052-5544-532	Sequence 522, Appl
C 33	71.8	0.8	11070	14	US-11-075-185-34	Sequence 34, Appl
C 34	71.8	0.8	78869	14	US-11-075-185-1	Sequence 1, Appl
C 35	71	0.8	1203	8	US-10-467-657-1859	Sequence 1859, Ap
C 36	71	0.8	1239	8	US-10-467-657-1857	Sequence 1857, Ap
C 37	70.2	0.8	3513	8	US-10-858-730-142	Sequence 142, Appl
C 38	70	0.8	8651	8	US-10-432-483-48	Sequence 48, Appl
C 39	69.8	0.8	3990	14	US-11-052-5544-520	Sequence 520, Appl
C 40	69.4	0.8	6360	14	US-11-056-470-2	Sequence 2, Appl
C 41	69.2	0.8	1590	8	US-10-858-730-162	Sequence 162, Appl
C 42	69.2	0.8	3240	14	US-11-052-5544-529	Sequence 529, Appl
C 43	69.2	0.8	4983	14	US-11-052-5544-531	Sequence 531, Appl
C 44	68.8	0.8	2562	14	US-11-052-5544-533	Sequence 533, Appl
C 45	68.6	0.8	2196	14	US-11-052-5544-539	Sequence 539, Appl

ALIGNMENTS

```

RESULT 1
US-10-506-454-1690/c
; Sequence 1690. Application US/10506454
; Publication No. US20060068386A1
GENERAL INFORMATION:
; APPLICANT: Slesarev, Alexi I
; APPLICANT: Mezhnevaya, Kat'ja V
; APPLICANT: Polushin, Nikolai N
; APPLICANT: Shcherbina, Olga V
; APPLICANT: Shakhova, Vera V
; APPLICANT: Malykh, Andrei G
; APPLICANT: Kozayavkin, Sergei A
; TITLE OF INVENTION: The Complete Genome and Protein Sequences of the Hyperthermophilic Bacterium Methanopyrus Kandleri AV19 and Monophyly of Archaeal Methanogen
; TITLE OF INVENTION: Methanopyrus Kandleri AV19 and Monophyly of Archaeal Methanogen
; TITLE OF INVENTION: and Methods of Use Thereof
FILE REFERENCE: FID001
; CURRENT APPLICATION NUMBER: US/10/506,454
; CURRENT FILING DATE: 2004-08-31
; PRIOR APPLICATION NUMBER: PCT/US03/06664
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: 60/361,742
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 1722
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1690
; LENGTH: 1694969
; TYPE: DNA
; ORGANISM: Methanopyrus kandleri
US-10-506-454-1690

```

Query Match	1.4%	Score 128.2;	DB 7;	Length 1694869;
Best Local Similarity	50.3%;	Pred. No. 6.7e-17;		
Matches 393;	Conservative 0;	Mismatches 358;	Indels 30;	Gaps 2

4198 TGATCAGTACTGGTAGTTGCTCTCCTCGTCCGGCCGGAGCGAGTGCACGGTGACCGCC 4257

Db 155792 TGACTGATGGTAAACGTGTTTCGCCCAGGGCTTCTCGGTGGTAGTTCGATGAGGGTT 155733

4258 GTACGTCGGCAGCTCGCTGTTAGAGCGCGTGGTTGACCCGGTTGTGCTCCCTGGTCT 4317

Db 155732 CTAATCGGCCAGTTCCTTGAGGTAGTAGCGGCTTCTCCCTCCTCCGCTCGTGAAC 155673

Qy	4318	CGCGGAACGCGTCCAGGGAGTGAAGCCCAATGGCCGCGGCGCACTCGCTCATTTGGCGGT	4377
Db	155672	CCTCCAGGCTCCGAAAGCTGGAAGATCCGATAGCGCGGCGAGTGTCCGTATCCGAAAGT	155613
Qy	4378	TGTGTGCGGATCTCGGTGACCACTTGTCCGGGCGGATGCGAAGTTGTGATGGCCGGA	4437
Db	155612	TGTACCTTAGCTTCACGTGATGTGACGTTCCGCGCTGCGCGTGTGAGCGAGATTCTGG	155553
Qy	4438	TCGGTTCGGCCAGGAGCGCGTGTGTGAGCAACGCGCCCGCTTCGAAAGCGGTGACCG	4497
Db	155492	CTTGTGTGCGGTGGAAGCTGAACACTCGGCGTTCACCGGATCCGCCACCGGACGTCCAC	4557
Qy	4558	CCGTGTGCAAGCCGAGGGCGTGGGGCGGCTGCAGAAAGAGCTTGACCTGGTGTCCGCGG	4617
Db	155432	TGTACTCGGCTCTGTGGGCGTGTGCGCGTCTTCGATCCAGATGAGATGTTCTTCGG	155373
Qy	4618	CGATCTTGCAGCGCCTTCACAGGTGCTGTGCGCCCAAGTGTACGCGCAGCATCG	4677
Db	155372	CGATCTTACAGATCCGGGTTCATGTCAAGCGGGGTGGCGTGAAGGTGCAAGGATCATG	155313
Qy	4678	CGCGCG-----TTCGGGGGTGACGAGCGCGGCACTGTT	4713
Db	155312	CTACCGTGCAGGAGCGTAGCACGCGGTCTTTGAGGAGCTTCAGACTCTTCAGAGGAGC	155253
Qy	4714	CCGGGTGCAGCAGAACCGGTGCGCGGCTGCAGCGTGCAGAAACACCGGTGAGTCCGAGCC	4773
Db	155252	GTGATTCATACGATGATGATGTTCCAGATGATGTCACGAAACGACGTCGCGCCCAAGT	155193
Qy	4774	AGCTGCGCGCGTGCAGGTGCGCGGCGGAGGTCAATTGACCGCATGATCACTTCAACG---	4829
Db	155132	GTAAGCGCGCGTTCGAGATGAGCATTAAGGTAAAGGTAAAGACGACGATGCGAGGTCCGCG	155133
Qy	4830	--GTACGTACCGGCGCGAGCACCAAGTTCCAGGGCGACGTTGCGGCTTTCAGGTGCGGA	4887
Db	155132	GCCCCAGCCCGCGGAGCTCCAGGGCGAGGTGCGGTGCGGCCGAGACTGTCGCGGA	155073
Qy	4888	TGCAGTCCGCTACCCCGACCAAGTCCGCGCACCCGCGCTCGAACTCCGACCAAGGCGCC	4947
Db	155072	CGCAAGTTCACATCCGAGCAACTGCGCAACTCAGCTCGAACTCCTCAACCGTGTGTC	155013
Qy	4948	C 4948	
Db	155012	C 155012	
RESULT 2			
US-11-143-980-1			
; Sequence 1, Application US/11143980			
; Publication No. US20050272133A1			
; GENERAL INFORMATION:			
; APPLICANT: He, Min			
; APPLICANT: Hucul, John			
; APPLICANT: Halcli, Bradley A.			
; APPLICANT: Magenat, Melissa M.			
; APPLICANT: Graziani, Edmund			
; APPLICANT: Sumner, Mia			
; APPLICANT: Kulowski, Kerry			
; APPLICANT: Pong, Kevin			
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex			
; TITLE OF INVENTION: Polyketide			
; FIDE REFERENCE: AM-101426US			
; CURRENT APPLICATION NUMBER: US/11/143,980			
; CURRENT FILING DATE: 2005-06-03			
; PRIOR APPLICATION NUMBER: US 60/664,483			
; PRIOR FILING DATE: 2005-03-23			
; PRIOR APPLICATION NUMBER: US 60/576,895			
; PRIOR FILING DATE: 2004-06-03			
; NUMBER OF SEQ ID NOS: 72			

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 116856
; TYPE: DNA
; ORGANISM: Streptomyces sp.
US-11-143-980-1

Query Match      1.4%; Score 122.8; DB 14; Length 116856;
Best Local Similarity 50.1%; Pred. No. 1e-15; Indels 39; Gaps 4;
Matches 443; Conservative 0; Mismatches 402;

Db      1259 CGGGAGATGATCCACAGATCGACCCGCGCGAGACACCGGCGCCTGTGCGAAGGTGTACACA 1318
Db      98628 CGGGCTTCTTCTTCCATGAGACCGCGCGAGACACACCGGCTACCGCGCACACTCACCG 98687
Qy      1319 GCGCTTACACCCCGCGTACGATCGCCGACCTTGAACCGCGCATCCGGAGGTGACCCGCT 1378
Db      98688 CCGAGTCTCGGTGCGCGCGCGACGCGAAGCTGACCGCGCGGATTCGAGCGGCTGCGCGAC 98747
Qy      1379 GCG-----TGTGTGCGGACCGCGCGGTGACCGCTTGAACCTGTGTGAGGCGCTCGCT 1429
Db      98748 GGCACCTCGATGCGATGCGATGAGAGCGCGCGGACGAGCGCGGACCTGTGTGCGGCTTACGCGCA 98807
Qy      1430 TCCCGCTGCGCGTCAAGATCGTGTGCGCGAGCTGTGTGCGGCGTGTGCGCGGATGACACCAAGC 1489
Db      98808 GTCCGGTGTCCCGGATGTGATGTCCGAATCTTGGCGTGTGCGGTGATACATACACACAGA 98867
Qy      1490 AGTTGCGTGACTGTGTCCGCGCGCCCTGTGTGACATCCAGATGAGACGACCCGACCGATCCGG 1549
Db      98868 AGTTGACACGAGGATACGACGCTCCGGGAGACCGCGCGGACGATCAGGCGCTGTGCGG 98927
Qy      1550 CCCGTGTGCAACGATCATGACAGGTGTGATACCCGCTACCTCTTACCTGTCTGACAGT 1609
Db      98928 CGATGCG-----GACCGCGGTGTGGAGCAAGATGCCCGGATTCG 98966
Qy      1610 GTCCGGAACGAGCGGCGCGACCCCGCGGAGACACTGATCTCCGCGCTGTGTGCGCGAGAG 1669
Db      98967 TGCCTGTCCAAACGAGCGCGACCCCGGAGACATGATCATGACAGGCTGTGTGATATGAG 99026
Qy      1670 TCGACGCGGCGCACCTTGCAGACGATGTGAGAGCGGCGCACTTCTCAACGCGTGTGCTCG 1729
Db      99027 TCGAGGCGGGTGTGCGTGAACCGAGAGAGATGTGTGCGCATGTGCGATCATCATTTTCG 99086
Qy      1730 CGGGGCAATCAACACCAACCGTCTGTGTGCGCAATGTCTCGGACCTCTGACGAGAC 1789
Db      99087 CCGGTCAATGAACCGGTGTGAGAACCTTATCGGCTGTGCGCATGTGTGCGCTGTTCCAGACG 99146
Qy      1790 CGAGGTCTGAGCGGCGCGCGCGAGAACCCCGGTGTGATTCGCGCGGATCATCGAGAG 1849
Db      99147 GTAGGACGTGACCCGGTGTGCGGAGAACCCGACCTTCAATGACGCGCGTGTGAGAGT 99206
Qy      1850 TGTGTGCG---TTTCCGCGCCCGCTTCCCGACAGTGTGACGCGACACGACGAGGCGACCA 1906
Db      99207 TCTTTCGTACTTCCCGGTGAACAATTTGTGGCACCGGTGCGACCGCGACGAGATGCG 99266
Qy      1907 CCGTGTGTGTGTGTGAGATCCGCGCGAGCTCATGTGTCAACACTGTGGTGTCTGTGCGCA 1966
Db      99267 TGATCAATGTGTACCCCATCGCAAGAGGCGAGATCGTGTGCGGTGTGTGTGTCAACCGCA 99326
Qy      1967 ACCGCAATCCCTGTGTGCGCATTCGGAACCGGACAGCTTGAACCGGTCCCGCAAGATCGGTG 2026
Db      99327 ACCGGAACCCCGAGCGGTGTGCGGATTCGACCGCTGTCTGTCAACCGGTGCGACA--- 99383
Qy      2027 GTGCGCGGAGACTTCTTGTGCGGACAGGCGGTGACATTTGTCTGTGTCGCGCGTGTGCGCG 2086
Db      99384 ---CTTCCACTGTGTGTGTGCGGACGAGTGTGACCAAGTGTGTGTGCGGCGGACGACTGTGCGA 99440
Qy      2087 GCTGTGAGAAACAGGTGCGCTGTGAGAGATCATGCGCGGTATC 2130
Db      99441 GGGTGAATGAGGTGTCTCTACAGGCGGTGTGTGTGAGTTTC 99484

```

US-11-075-185-53
; Sequence 53, Application US/11075185
; Publication No. US2005026643A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; PRIORITY FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 1221
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-11-075-185-53

Query Match 1.4%; Score 121.6; DB 14; Length 1221;
Best Local Similarity 47.7%; Pred. No. 2,4e-15;
Matches 500; Conservative 0; Mismatches 524; Indels 24; Gaps 4;

QY 1191 GGTCTCTCCGAGACACCCCTTCTCTCCGACCCCGGTCTATCGAGGGGCGCA 1250
DB 189 GGGGCTCTCTCGGGCTCTCGGCAATGCTCGGAGAGACGACCTCTCGAGAAAGCTTCA 248
QY 1251 CCCGACCGCGGGGATGATTCACGAGATTCACCGCGGAGACACCGGCGCTTGGGCAAGT 1310
DB 249 CTCCTGTGCGGCTGCGCTTCTTCCAGAGACCGGCGGCAACGAGCTTGGACGCT 308
QY 1311 CGTACAGAGCGCTTCAACCCCGGTACGATGCGGACCTTGAACCGGCAATCCGAGAGT 1370
DB 309 CATCATGAAGGGTTCTGCGCTTCTCGGCGGTGAGTGGATGCGCCGCGGCTCGTGGGGCT 368
QY 1371 GACCCGGTGTG-----CTGCTGCGGCGAGCGCGGTGACCGCTTGAACCTGATCGAGCGCT 1424
DB 369 CGTACAGCGGCGCATGAGGGGGCGGCGCGGACGCGAGATGATGTCCTCTCGGCGTT 428
QY 1425 CGCCTTCCGCTCCGGTTCAGATTCGCGGAGCTGCTGGGGCTGCGCGGATGAGCA 1484
DB 429 CTCGAGGGGCTGCGCTTCAACGCTGCGGAGATGTTCTGATACCGAGGTGATCG 488
QY 1485 CAAGCATTCGATGCTGTCGCGCGCTGCTGATTCATTCAGATGAGACCGGACCGA 1544
DB 489 CCGCAGTTCTGAGATGCTGACCGATCTCTTGAAGCTGCGCGGCGGCGGTGAGCTC 548
QY 1545 TCCGCGCTGCTGTAACGATTCATGAGTGTGTAACCGGCTCACTCTGCTCGA 1604
DB 549 GAGGAGCAGAGACCGGCGGCTGAGAGAGCTCTCGACATCTCGACTGATGATGAG 608
QY 1605 CAGGTTCGGGAGACCGGCGGCGGACCCCGGAGACGATTCCTCCGGTGTGCTGGC 1664
DB 609 GCTCTGTCAGAGACCGGAGAGCGCGCGGAGAGAC---GTGCGAGACAGTTCACTCG 665
QY 1665 CGAGGTTCAGAGCGGCGGACCTTCAGACGATGAGAGCGGCACTTCTTCAACAGTTGCT 1724
DB 666 GGGGAGAGACGATGAGACGAGAGCTCGCGGAGAGGCGGCAATGATGCTTCCAGATGCT 725
QY 1725 GCTTCGCGGAGCATCAACACGCTCTGCTGGGCAACATGTCGACCTTCGACGA 1784
DB 726 CGCGCGCGGATTCGATCACTCTCTGTAACGATTCGGAACACCGTCTCGGCTCTCAA 785
QY 1785 GACCGCGGAGTATGAGACGCGCGCGCGGAGACCGGCGGTCTGATGCGCGGATTCATCA 1844
DB 786 CACACCGCGAGACCTGCGAGAGCTGCGGAGAGCGCGGCGCTGTCGCGCGCGGTCA 845
QY 1845 GAGAGTGTGCGTTTCGCGCGCGCGGTTCCTCCGATGACGCGGACGACGAGGCGCAC 1904

DB 846 GAGAGCTGCGCTTCGAGCGGCTCGGTCTCTCTTCAAGCGGATGAGAAAGACAC 905
QY 1905 CACCGTGTGAGGATGAGATCCCGGCGGACGATGATGATCAACCTGGGTGCTTGGC 1964
DB 906 GAGATTCGGGGGCGGAGGATGTCGAGAGGCGAGTTGCTTTCGATGATGATCCGAGC 965
QY 1965 CAACCGGATTCCTGAGCGATTCGAGACCGGAGACAGTTGACCCGTCGCAAGATCGG 2024
DB 966 GAAACCGGATTCGAGGCTGTTCTTCCAGCGGATTCGATGATGATCAACCGGAGCAGAG 1025
QY 2025 TGGTTCGCGGACCTTCTTCGAGGACGCGGTGACCTTCTGTCGTGTCGTCGCGCTGAGC 2084
DB 1026 -----CCGACACTGACTTCGAGGCGGCGGTGATGACTGCGCGGAGCGCGCTCAT 1079
QY 2085 GCGCTGAGAGACGAGTGGCTTGAAGAGATGATGCGCGGTGAGCTGAGTGGCGCT 2144
DB 1080 CCGGATGGA-----AGTAGAGAGTGCCTGCGCGCTCTCTCTGCTGCGCGCTG 1130
QY 2145 CGACCGGAGACGACGCTGCTGATCTTTCAGACGATGCTGCGGACCGCGGACCT 2204
DB 1131 GAGGCTGCGGAGAGACGTTGATGCTACCGCGGCTGAACTGACAGACCGCGGCGAG 1190
QY 2205 CCGGATGTCGCGGCGGTCAACCGCGGC 2232
DB 1191 CTCGCTGCGGCTTCTGCTTCCCGGACGC 1218

RESULT 4
US-11-075-185-2
; Sequence 2, Application US/11075185
; Publication No. US2005026643A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; PRIORITY FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 14172
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-11-075-185-2

Query Match 1.4%; Score 121.6; DB 14; Length 14172;
Best Local Similarity 47.7%; Pred. No. 2e-15;
Matches 500; Conservative 0; Mismatches 524; Indels 24; Gaps 4;

QY 1191 GGTCTCTCCGAGACACCCCTTCTCTCCGACCCCGGTCTATCGAGGGGCGCA 1250
DB 5329 GCGCTCTCTCGGCGGCTTTCGATGCTCGGAGAGACGACTTTCAGAGAAAGCTTCA 5388
QY 1251 CCCGACCGCGGGATGATTCACGAGATTCACCGCGGAGACCGGCGCTTGGGCAAGT 1310
DB 5389 CTCCTGTGCGGCTGCGCTTCTTCCAGAGACCGGCGGCAACGAGCTTTCAGAGCT 5448
QY 1311 CGTACAGAGCGCTTCAACCCCGGTACGATTCGCGGACCTTGAACCGGCGATTCGGAGGT 1370
DB 5449 CATCATGAAGGGTTCTGCGCTTCGCGCGTGAAGTGAAGTGCCTGCTGAGGCT 5508
QY 1371 GACCCGGTGTG-----CTGCTGCGGAGCGCGGTGACCGCTTTCAGACTGATGAGCGCT 1424
DB 5509 CGTACAGCGGCGCATGAGAGGGGCGCGCGGACGCGGAGATGATGCTCTGCGCGTT 5568
QY 1425 CGCCTTCCGCTGCGGCTTCAACGATGCTGCGCGGAGCTGCTGAGGCTGCGCGGATGAGCA 1484

Db	5569	CTCCGAGGACGGTTCGCGCTCAACACGCTGCGCGAATGTTTGATGATACCCGAGTGCATCG	5628
Qy	1485	CAAGCAGTTCCGATGACTGTGTCGGCGCCCTGATGCAATCCAGATGACGACCCGAC	1544
Db	5629	CCCGCAGATTCTTGAGATGATGTGACCCGATCTCTTGAACTCGCCGGCGCGGGATGAACTTC	5688
Qy	1545	TCGCGCCCTGGTCCGAACGATCATGACGATGCTGAACCCGCTGACCTCTTACCTGCTCGA	1604
Db	5689	GGAGAGACGAGAGCGCGCGGATGAGAGAGAGCTGCTGCGACATGCTTCGACATGATGAG	5748
Qy	1605	CAGGTGTCCGAAAACGCGCGGCGCGAACCCCGGAGACGACTGATCTCCCGGCTGTGTGCGC	1664
Db	5749	GCTGTGCGAGGAGCGCGCCGGAAGCGCGCGGGGAGAGAC---GTCCGAGACAGATTCACTGC	5805
Qy	1665	CGAGGTGACCGGCGCGCACCTTCGACGACGTGAGAGCGCGCACTTCTTCACAGCGTTGCT	1724
Db	5806	GGCGGAGGACGCGTGAACACCGAGCTTCGCGGCGGAGGCGCGCCATGCGAGTCTTCAGATGAT	5865
Qy	1725	GCTGCGCGGGGACATCACACCAACCGTCTCTGTGGGAAATGCTGCGCAACCTTCGACGA	1784
Db	5866	CGCGCGCCGGAATTGTCACCTCCGTGAAACAAATGCGCAACACGATGCTCGGCTCTTCAA	5925
Qy	1785	GCACCCGGAATACGAGCGGCGCGCGCGCGAGAGAACCCCGGATCTGATATCGCGCGCATATCGA	1844
Db	5926	CCACCCCGCAAGCTCGCGGAAGCTGCGGAGAGCGCGCGGCTCTGTCGCGCGCGCGGTGCA	5985
Qy	1845	GGAAGTGTTCGTTTCCGCGCCCGCGTTCCCGCAATGCGACGACGACGAGGAGCCAC	1904
Db	5986	GGAAGGCTCGCGCTTCGAGCGGTCCGATCTCTCCCTCAGCGCGCATGTGCAAGAGAGACAC	6045
Qy	1905	CACGATCGATGGGGGTTCAGATATCCGCGCGCGACGTCAATGTTCAACACTGTGGGTGCTCTCGC	1964
Db	6046	CGAGATCCGCGGCGCGCGCAAGGATGTCGAGAGGAGCATTTGTTTCGAGATGATCGCGGACG	6105
Qy	1965	CAACCGCGATCCCGTGGGCGCATCCCGGACCCCGGACAGTTTCGACCCCGTCCCGCAAGATCGG	2024
Db	6106	GAACTCGCATCCCGGGCTGTTCTCCGAGCCGATGATTCGATATCACCCCGGACGAGAG	6165
Qy	2025	TGTGTGCGCGAGCTCTCTTCGCGGACGCGGTGACATTTGTTCTCGGTGCTCCGCTGGC	2084
Db	6166	-----CGGGAACCTGACCTTCGCGGAGCGCGCTCATTAATGACC CGGAGGCGCCGCTCAT	6219
Qy	2085	GCGCTTGAAGAACAGATGCGCTTGGAGAGATCATGCGCCGCGATCGGTCACTGGCCGT	2144
Db	6220	CCGAGATGAA-----AGTAGAGAGATGCTGTGCGCGCCCTGCTCTGCTGCGCGCTG	6270
Qy	2145	CGACCGGACGACGACGACGCGTGCATCTTCGACCAAGATGCGTCTCGGACCCCGGACCT	2204
Db	6271	GGAAGTCCCGCAAGAGAGATGTGAGCTTACGCGCGGATCGAACTTCGAGAGACCGCGGCGGAG	6330
Qy	2205	CCCGGTGCTGCGCGCGGTCAACCCCGGCGC	2232
Db	6331	CTCGCTGCGCGCTTTCGCTTCCCGGAGGCC	6358
RESULT 5			
US-10-497-135-17			
/ Sequence 17, Application US/10497135			
/ Publication No. US20050272132A1			
/ GENERAL INFORMATION:			
/ APPLICANT: Biologica Technology Ltd			
/ APPLICANT: Gregory, Matthew A			
/ APPLICANT: Galsner, Sabine			
/ APPLICANT: Petkovic, Hrvoje			
/ APPLICANT: Moss, Steven			
/ TITLE OF INVENTION: Production of Polyketides and Other Natural Products			
/ FILE REFERENCE: 4408-P0344US00			
/ CURRENT APPLICATION NUMBER: US/10/497.135			
/ CURRENT FILING DATE: 2004-05-26			
/ PRIOR APPLICATION NUMBER: PCT/GB03/003230			
/ PRIOR FILING DATE: 2003-07-16			
/ PRIOR APPLICATION NUMBER: GB0216509.0			
/ PRIOR FILING DATE: 2002-07-16			

```

? PRIOR APPLICATION NUMBER: GB0224922.5
?
? PRIOR FILING DATE: 2002-10-25
?
? NUMBER OF SEQ ID NOS: 63
?
? SOFTWARE: PatentIn version 3.2
?
? SEQ ID NO 1
?
? LENGTH: 767
?
? TYPE: DNA
?
? ORGANISM: Streptomyces hygroscopicus
?
? US-10-497-135-17

```

Query Match	Score	DB	Length
1.3%	118	8	767

Matches 351; Conservative 0; Mismatches 340; Indels 9; Gaps 2

1459 CTGCTGGGCTGCCCGGATGGACCAAGCAGTTCGGTGA CTGGTCCGGCGCCCTGCTC 1518

b 6 CTGGCGTCCCTGGCCATCCACGACCTCTACGGCCTGATGAGGAGGGGCCCTACTC 65

1519 GACATCCAGATGACGACCCGACCGATCCGCCCTGTGGAACGCATCATGCAAGTGTG 1578

66 GAGGCCAGATGCCGGCCATGAGGGCGGCACCGACATGAGAGCATCAAGAGGCTGACC 125

1579 AACCCGCTCACCCTACCTGCTCGACAGGTGTGGGAACGGCCGGCCGACCCCGGAC 1638

126 GACGAATCTTCGGTCACTCTGGCGCTGGTGCCTGCCAAGCGGAGCAGGCGGCGAC 185

1639 GACCTGATCTCCCGGCTGTGCTGGCCGAGTGCACGGGGCACCCTCGACGACGTGAG 1698

b
186 AGGCTTCTGCACCGGCTGGCCGAGTCCGGCGAGGACGAGATCCTGCTCAGCGACGAGAG 245

1699 GCGGCCACTTCTCCACAGCGTTGCTCGCGGGCACATCACCAACCCTCTGCTG 1758

246 GCGACCGGGTGTTCGCCACTCTGCTGTTGCCCGGGCAGCACTCGATGCAGCAGATGGTC 305

1759 GGCACATCGTCCGCACCCCTCGACGAGCACCCGGAGTACTGACGGCCGCCCGAGGAC 1818

b 306 GGCTACTGTCTGTACGGCTGCTCTCCATCCCGAGCAGCGGGCGGCGCTGCCGGAGAAC 365

1819 CCGGCTGATCGCGCGATCATCGAGGAGGTGTGCGTTTCCGCCCCCCCGTTC--CCC 1872

b
366 CCGA**C**CTGATCGACGGCGGT**G**CGAGAGCTGCTGC**G**CTTCC**T**GCCGCTCAACCAGCTC 425

1876 CAGATGCAGCGCACCAACGACGAGGCCACCGTCGGTGGGTGAGATCCGGCCGAC 1933

b 426 GGGCTGCCGGGTCTGTGTCGAGGACGTGAGCTGCACGGCCAGACCATCAGCGCCGGC 485

1936 GTCATGTCACACCTGGTGCTCTCGGCCAACCGGATCCCTGGCGCATCCCGACCCG 1993

486 GACAACGTGATCCCGCTCTACTGCACGGCCAAACCGCGACCCCGCGTCTTGGCCGACCCC 545

1996 GACACGTTGACCCGTCGCCGACAGATCGGTGGTGCCGGCGCAGCTCTCCTTCGGGCAACGGC 205:

546 GACACGTTGACATCAGCGTAAGCCGGAACACAACT-----TCGCTTTCGGGTACGGC 599

2056 GTGCACTTCTGTCTCGTGCCCCGCTGGCGCGCTGGAGAACGAGTCCCTCGAGGAG 2111

600 ATCCACAAGTGCCCGGGGAGCACCTCGCCCGCGTGTGATCAAGTCCCAACGCTGCC 659

2116 ATCATCGCCCGGTACGGTGGCGTGCACCGCGACG 2155

660 CTGTTGAGCGCTTCCCGGATGTGCGACTGGCGGGCGACG 699

SECRET

US-11-143-401-74

Publication No. US20060002953A1
CURRENT INFORMATION

APPLICANT: Behr, Marcel
ADDRESS: 52311 Dettm

APPLICANT: Schoolnk, Gary

TITLE OF INVENTION: Molecular Differences Between Species of

```

; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/11/143,401
; CURRENT FILING DATE: 2005-06-01
; PRIOR APPLICATION NUMBER: US/10/647,089
; PRIOR FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US/09/894,844
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FaSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-11-143-401-74

```

Query Match 1.3%; Score 115.6; DB 14; Length 1215;

Best Local Similarity 49.5%; Pred. No. 4e-14; Indels 27; Gaps 4;

```

Matches 428; Conservative 0; Mismatches 409; Indels 27; Gaps 4;
QY 1279 GACCCGCGGAGCACCAGGCTCTGCGCAAGTGTGTCAGACGCGCTTCAACCCGCGTACG 1338
DB 277 GATCCGCGGTCACACCGAGTTTCCGACCTGTGTGCGGCGCTTCAACCCGCGTACG 1336
QY 1339 ATGCGGACCTGCAACCGGCAATCCGAGAGTGAACCGGTGCTGTGCGCA-----C 1392
DB 337 GTGAAACCTTCAGGCCCAAGTGTGCGCAAGTGTGCGGTGCGGTGCGCAAAAGTGTGCG 396
QY 1393 GCGGAGTACCGCTTTCAGCTGTGTGAGGCGGCTGCTTCCGCTGCGGCTGCAAGTGTGTC 1452
DB 397 GCCAAGGTGTGCGGCGGCAATTGTCAAGAACTAATTCAAACCGCTGCGTGTGTGTG 456
QY 1453 GCGGAGTGTGCGGCTGCGGCGGAGTGAACCAAGAGTGTGCTGTGCTGTGCGGCGC 1512
DB 457 GCGCACTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 516
QY 1513 CTGTGTGACATTCAGATGAGCAGACCGGATCCGCGCTGTGTGCAAGCATGATGACG 1572
DB 517 ATGTGTGCGGCGAACCAGGTTGACGCGGCGCACACG-----GCGCATGTGAC 564
QY 1573 GTGTGTGACCGGCTCACTCTCTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1632
DB 565 GCGGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 624
QY 1633 CGGAGCAGCTGATCTCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1686
DB 625 GCGGACGAGCGCATCTTCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 684
QY 1687 GACGACGTGAGGCGGCGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1746
DB 685 GGCACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 744
QY 1747 ACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1806
DB 745 ACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 804
QY 1807 GCGCGGAGGAGCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1866
DB 805 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 864
QY 1867 CCGTTTCCCCAGATGAGCGGACACGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1926
DB 865 CCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 924
QY 1927 CCGGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1986
DB 925 CCGGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 981
QY 1987 CCGGACCGGAGCAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2046

```

```

DB 982 TACGGCGCGGAGCGGAGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1041
QY 2047 GGGCAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2106
DB 1042 AGCCACGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1101
QY 2107 CTGAGGAGATGATCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2130
DB 1102 CTGACCGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1125

```

RESULT 7

```

US-10-497-135-18
; Sequence 18, Application US/10497135
; Publication No. US2005027132A1
; GENERAL INFORMATION:
; APPLICANT: Biotech Technology Ltd
; APPLICANT: Gregory, Matthew A
; APPLICANT: Galsner, Sabine
; APPLICANT: Petrovic, Hrvoje
; APPLICANT: Moss, Steven
; TITLE OF INVENTION: Production of Polyketides and Other Natural Products
; FILE REFERENCE: 4408-P0344US00
; CURRENT APPLICATION NUMBER: US/10/497,135
; PRIOR FILING DATE: 2004-05-26
; PRIOR APPLICATION NUMBER: PCT/GB03/003230
; PRIOR FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: GB0216509.0
; PRIOR FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: GB0224922.5
; PRIOR FILING DATE: 2002-10-25
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 761
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-497-135-18

```

Query Match 1.3%; Score 112.2; DB 8; Length 761;

Best Local Similarity 51.8%; Pred. No. 2e-13;

Matches 308; Conservative 0; Mismatches 278; Indels 9; Gaps 2;

```

QY 1564 ATCATCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1623
DB 105 AGCATCAAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 164
QY 1624 GCGGACCGCGGAGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1683
DB 165 GAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 224
QY 1684 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1743
DB 225 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 284
QY 1744 ACCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1803
DB 285 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 344
QY 1804 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1863
DB 345 GCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 404
QY 1864 CCGCGTTTCCCCAG---ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
DB 405 CCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 464
QY 1921 GAGATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
DB 465 ACCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 524
QY 1981 GCGCATTCGAGCAGCGGAGCAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040

```

Db 525 GTCCTGCGCAGCCCGACAGCTTCAGCTACGCGGTAAAGCCC-----GAAACCACTTC 578
 Qy 2041 TCTTCGCGGACGCGCGTCACCTCTGTCTCGATGCCCGGCTGCGCGCTTGAGAAACCG 2100
 Db 579 GCTTTCGGGTACGGGATCCACGCGCTGCCGCGGAGCAGCACTCGCCCGCGTGTATCAAG 638
 Qy 2101 GTTCGCCCTGAGAGAAATCATTCGCCCGGATCGAGGTGCATCTGCGCGTTCGACCCGGAGG 2155
 Db 639 GTGCGCACCTGCGCGCTGTTCGAGCGGCTCTCCCGAATGTCGCACTGCGCGAGCGAGG 693

RESULT 8

```

US-11-143-401-127/c
; Sequence 127, Application US/11143401
; Publication No. US20060002953A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/11/143,401
; CURRENT FILING DATE: 2005-06-01
; PRIOR APPLICATION NUMBER: US/10/647,089
; PRIOR FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US/09/894,844
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-11-143-401-127

```

Query Match	1.2%;	Score 107.8;	DB 14;	Length 1236;
Best Local Similarity	49.5%;	Pred. No. 1.5e-12;		
Matches 410;	Conservative	0;	Mismatches 32;	Indels 27;
			Gaps	4
Qy	4261	CGTCGGCAGCTCGCTGCTGTGTAGAGCGCGTGTGACCCGGATGTGCTCCTGATCTCGG	4322	CGTGTGCTCCTGATCTCGG
Db	919	CGTCGGCCATACCGGTGCGATAGCACTCGAGAGACCTTGGGGGCACTTGTCCAGCGGCAT	860	
Qy	4321	CGAACGCGTCGAGGGAGTGAGCGCCCATGCGCGGCGCATTCGCTCATCTTGGCGTTGG	4381	
Db	859	CAAGCCGCACTGTTGGCGTAGGCCAATAGCGCGCTGATCTCCGACACACTTCGCGTTCA	800	
Qy	4381	TGCGCATCTCCGAGACCACTTGTGCCGGGCGCATGCGAGTTGTGTGACATGGCCGGATCC	4444	
Db	799	TTCCGCGCTGATGTGACTGCGCGTGTTCACCAAGCGAAGTTCTGGAATTGTATGCT	740	
Qy	4441	GTTCCGCGCAGACGAGCGCGTCTGTCGTGACCAACGCGCCCGCGCTCGAAGCGATGACCGCT	4500	
Db	739	GCCTCGACGAGCGGTGATCGCGAGAAACCAAGACCGCGCGCTTCACCAACCGGAAACGGCT	680	
Qy	4501	TGCTGCGCGTGAAGACTGAACACTTCGGCGCTCACCCGATTCGGCCACCGGACGTCCACCG	4561	
Db	679	TGCTTCACATGAAGAGGAAGATCTCCCATGACACACCGGTCACACGAGCGCTCGCGCTCG	620	
Qy	4561	---TCGTGACGCCCAAGGCGTGGCGCGCGCTCGAAGAAAGACTTGAACCTGTGGTGGACGG	4611	
Db	619	CGTACGTGAGCGGAAGCCGGCGCGCGGATCGAGACCAATCGGTACCTCCCATTCGGCGG	560	
Qy	4618	CGATCTTC--GCACGCGCTTCACAGGTGCTGTGGCGCC-----ACAGTGTGA	4661	
Db	559	CGAGCTCTCTCCCGAGACGCTGATCTGGGAGATTGCGACGCGCGAACAATTGGCCAGACGGA	500	

OY	4666	CGCCGACGATCGCGCGCGTCTGCGCGGGTGA	CGACGCTGGCGAGCTGGTCCGGGTGCACCA	4725
Db	499	TGCGCGGATCCGGTCCCGAAGCGTTGATGA	CGCGCGCGGGCGGAATGACGATGGCT	440
OY	4726	GACCGGTGCGCGGGTGCACGTCGCGAAGCA	CCGGTGTAGTCCGACGACGCTGCGCGGT	4785
Db	439	GCCATGTGTGGCGCTGCATGTGCATGA	CCAGGACGGTACCCAGTCCATAGCGACGCT	380
OY	4786	GCGCGGTGCGCGCGAAGTCATTGA	CGGCAATGATCACTTC-----ACCGGTGA	4833
Db	379	GAGCGACGCGGACGAACTGAACGA	CGGCATCAGCAGTAGTACGGTCCCGCGTACCGCGCG	320
OY	4834	CGTCAACCGGCGCGCGACCA	CAGTTTCCAGGGCGACAGTGGCGTTGCACAGTGGCAATGCAGT	4893
Db	319	CGAAATCTGACCTGTAGCGCGCGCGAGAGTGC	CAGGGTGCCTGTGGCCAGGGGTAGCAACGT	260
OY	4894	GCCGTATCCCGGACCGAGGTGCGCGAC	CCCGGGCTCGAACTCCCGCACGAGGAGCCCGCGGT	4953
Db	259	GCAAGTAGAGTCCCAAGATAGTTCGCGC	AGGGCGCGGCGCAAACCGCGCTCTGTTCCGAACGA	200
OY	4954	TGTGTAGCCAGTTGTGTTTCAGGGCCCA	CTTCAGGCGGGGCCAAGAACCGCTGCGCGTGC	5013
Db	199	AGTTGTGTATCAGATTAGCTTGGGCGAGAT	CTGTATCGAAGTCTCTCGGCGAGCTCGGCTGACC	140
OY	5014	CGATGTCGCGCGGCGCCACGTCGAC	AGGGGGGTGCAGGAAACGCTCGGGAAC	5062
Db	139	CGGGAAGACTCGGGCGAGTGA	AGGGGAACTTTGGGAACTGTGCAGCACCC	91

RESULT 9

```

US-11-096-568A-20202/c
; Sequence 20202, Application US/11096568A
; Publication No. US20060048240A1
;
GENERAL INFORMATION:
;
APPLICANT: Alexandrov, Nikolai et al.
;
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Poly
;
TITLE OF INVENTION: Therdly
;
FILE REFERENCE: 2/50-1592PUS2
;
CURRENT APPLICATION NUMBER: US/11/096,568A
;
CURRENT FILING DATE: 2005-04-01
;
NUMBER OF SEQ ID NOS: 34471
;
SEQ ID NO 20202
;
LENGTH: 1884
;
TYPE: DNA
;
ORGANISM: Zea mays subsp. mays
;
FEATURE:
;
NAME/KEY: misc_feature
;
LOCATION: (1)..(1884)
;
OTHER INFORMATION: Ceres Seq. ID no. 12379414
;
US-11-096-568A-20202

```

Query Match	Similarity	Score	DB	Length
Best Local	44.0%	Pred. 1.7e-09		
Matches	447	Conservative	0	Mismatches 535; Indels 9; Gaps 1;
QY	2570	GGGATGACCCCGGCGACGTAAGCGGTGCGCCCGGCGACGCCGGGAAAACCGGCTCC	26228	
Db	1130	GAGGCGAGGTGAACGGCGCCGAGATGACGGCGCCCATGACCGCGGCTTCTCGACAA	10778	
QY	2630	CGGTACAGGTAGACTCTCCGACGACGCTGATCTGACCGGCACTCGGAGATGGGCGGTC	26885	
Db	1070	CCGGCGCTCTGGAACGACACGGGCGAGAAAGCGAGTACATGATACCCCGGTAGCTGC	10111	
QY	2690	GGGCGCATCGTTTCCGGCCGGAATCCGCAACAGCTGGGCGTGCACACCCCGACGACGCTC	27488	
Db	1010	TGGAAACAGGGGACCGGCACCGGCACCGGCACAGTGTATCCGGTGCTCCCGCGGAGATC	951	
QY	2750	TCCAACGCGTAAACCGAGGTCGGTCTGCAATGCCCCGGGGTCCGCTCGGCGCGATGCGAGC	28085	
Db	950	CTGCGGAACGCGCCCTGCTCTGTGGGCGGCGCGGCTTCACAGCGCGGGCACGCTCTCC	891	
QY	2810	AACCGGAGAGGCGCTCCCGCAACTCGGTCTGCTTCCCTCGAACACTGCGCTGCTCC	28658	


```
FEATURE:
NAME/KEY: misc feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1
```

Query Match 1.0%; Score 92.6; DB 14; Length 88421;
Best Local Similarity 42.4%; Pred. No. 1.3e-09;
Matches 1317; Conservative 0; Mismatches 1729; Indels 57; Gaps 13;

```
QY 2391 CGGCGCGATCAGCGCCGATCTGATCCACCGTCGCGCGACGTCCTCGTTGAGTTGAT 2450
DB 47486 CGACGTATATCTCGGCGAGGCGCGGCGCTCTTACCGGCGCGAGATGAGCCGCG 47427
QY 2451 TGAGCGGCGCGGTCAACGACTGTGAAACCGTCGAGGAGAACTGTCCCGCGCTCG 2510
DB 47426 TCTCTCGGCTTCAGCGGCGGTGACCTCGGCGAGCGCGCGGATCGCGCGAGACA 47367
```

```
QY 2511 GATCGATGCTCGGCGCGACGTGACGAAAGTCGTGACGACCGAGTGACGCTCCGCTCG 2570
DB 47366 CGCGAGAAACGCGGCTTCAAGTACCGCGGCGAGCGCGGCGCGTGGCCGCTCGAAGAGGT 47307
QY 2571 GGGTGAACCGCGCGCGACGTAAGCGGGTGCCCGCCGACCGCGGAGAAACCGGCTCC 2630
DB 47306 CGGCGCGCGAGTGAAGTTCACCGCGGCGAGCGCGCGCGCTCCCGCGCGCTCGAAGGCT 47247
QY 2631 GGTACAGTGAAGTCTCCGAGAGGTCATCTGACCGCACCTGCGAGATGAGCGGCTCG 2690
DB 47246 CGGTGAACGTGAGTTCAGGTTCGAACTTGCCCGAAGCGGCGCGCTGCGACACGACA 47187
QY 2691 GGGCATCTGTTCCGCGCGGATCCGCAACGACTGCGGTGACACCCCGACGAGCTCT 2750
DB 47186 TCGCGAGCGCGGAGCTCGGCGCGCGTCCGCGCGCGCGCTCGAGGATGACATGACCT 47127
QY 2751 CCAACGCGTAAACCGAGTCCGCTGATGATGATCCCGGCGTCCGCTCGCGCGGTGACGA 2810
DB 47126 GGAACAGGCGGTGCGGCGCGAGCGAGCGGCGCGGCTGAGCTCTTCGACGACTTCTCGA 47067
QY 2811 ACCGCGGAGAGCCCTCCCGAATCGGTCGCTCGCTCGGACCAACTGCGCGCTGCC 2870
DB 47066 ACCGAGCTCTGAGTGGCGCGAGCGCCGCAAGGTCAGCTGCGCACCTTGCGCG-----A 47013
QY 2871 GACCGCTGATGCTCTCGCGACGAGTGAAGAGTCGAGCGCGCGGTGCGGCGCTGACTCGT 2930
DB 47012 GCACTGCTGTCAGCGCGCGGCTCGCGGCTCGAGGTGAGTGGCGAGACGAGCGTGTGACGA 46953
QY 2931 TGAAGTCGAGGATGATGACGACGAGTGAAGAGCGGTTGCGCGCGCGCGCGCGAGTGA 2990
DB 46952 AGAAGCGGACGAGGTTCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46893
QY 2991 TGTGTTGAGGCGGAGCGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3050
DB 46892 GAGCTGAGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46833
QY 3051 GGAAGTCTTCTGTAACCGCTTCGAAAGCGCGCGCGCTTGCAGTGGGCTGTGATCTGT 3110
DB 46832 TGAAGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46773
QY 3111 CGTTGTTCAAGCGCGTGAAGAGGATGAGACCGCGCGCGAGTCCGACGCGCGGCTGCG 3170
DB 46772 GGAATCTCGCGCTGTGAGGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46713
QY 3171 GGGCGACGCTGTTTCCGTGAGGGCGGAAGCGTGTGTGAGACGCTGAGGGCGGAAGCGCC 3230
DB 46712 CGGCGAGGTCAGGCTCTCGCGCACGCGCTGAGCGCGCTCGCGCAAGTGGCGATCTGCG 46653
QY 3231 GCCGCGCGCGTGCAGACCAACGCGCGCGATACCGCGGTTGATCAAGTTCAGAGCGCG 3290
DB 46652 GGGAGATCAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46593
QY 3291 CGGAGAGTACATCGCGCTGCGGTTTCCGA--CGGACCTGTCGATGACCGAGGTGAG 3348
DB 46592 AGTGGGTACTGACGCGCGAGGCTTCGAGCTCGGCGCGCGCTGCGCGCGCGCGCGCG 46533
QY 3349 GATCTCGTTCGCGCGCACACCGACTCGCGGTGATCTTGCGACCTGACCGGTAACGA 3408
DB 46532 CGTGAAGCGGTGCGAGGTCGCGGCGCGAGCGCGCGATGACACGCGTGGCGCGATGT 46473
QY 3409 GAAATGACAGCGGAAATGACAGTCCG-----GCCGAGGTAGAGCCGAGCTGTAGC 3461
DB 46472 GGTGACAGCACGACGACGACGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 46413
QY 3462 GAAAGCGCGGCTCCCGCGTACCGGAGCTTCAACGCGCGCGCGCGCTTACGCGGCTGGA 3521
DB 46412 GCGGAGATCTCCCGGCGCGGTCGAAAGGCTGACGCGCGCTTCCGCGCGCGCGCTCTTCGA 46353
QY 3522 TGTGTTGCTCAATGATCTTGCGCGCGCGCGCGCTGCTCAACCGCGCGCGAGCTGCGGGA 3581
DB 46352 CCGCGCGCGCGCGAGCGTCAACGACCGGCAACGCGGCTTCCGCGCGCGCGCGGAGTGAAGA 46293
```

QY 3582 TGGGATGAAACAGGTGAGCAGCTGCGCAACCCGCGCGGTTCAGGTGCAACCGACGCGC 3641
 Db 46292 TGTGTTCACACGGCTTCGCGCGGTTCGCGCAACCGTTCAGGACGCTCGTGGCGG 46233
 QY 3642 GCATCTCTCACAACGGGTGTAACGGCAGCGCGGTGATGGGCGGTGATTTGCGCAGCC 3701
 Db 46232 CCACACGTGCGCGAGGGCGCGGTGATGGGCGGTGATTTGCGCAGCC 46173
 QY 3702 GGAACCGCTGTCTGTTGCGCGGTGATGGGCGGTGATTTGCGCAGCC 3756
 Db 46172 GCGGACCGGGATTTTGTAGTGGCTCGGTCTTCAAGTGGGCAAGAACCAAGCC 46113
 QY 3757 GGGCGTGAAGCGCGCGCAACCGCGCGCGGTGATGGGCGGTGATTTGCGCAGCC 3816
 Db 46112 GGGCGTGAAGCGCGCGCAACCGCGCGCGGTGATGGGCGGTGATTTGCGCAGCC 46053
 QY 3817 CTCCGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 3876
 Db 46052 GGTGTGTCGACCGCGGTGATGGGCGGTGATTTGCGCAGCC 45993
 QY 3877 GTCGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 3932
 Db 45992 CGGACGACCGAGACCGCGGTGATGGGCGGTGATTTGCGCAGCC 45933
 QY 3933 --ATCAACGACCGCGGTGATGGGCGGTGATTTGCGCAGCC 3990
 Db 45932 GCGACGAGTGCCTCGCGCGGTGATGGGCGGTGATTTGCGCAGCC 45873
 QY 3991 CACTGACACCGAGCGCGGTGATGGGCGGTGATTTGCGCAGCC 4041
 Db 45872 GCACTGCGAGAGCGCGGTGATGGGCGGTGATTTGCGCAGCC 45813
 QY 4042 ACTGCGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4101
 Db 45812 CACCGGCTTGAAACCGCGGTGATGGGCGGTGATTTGCGCAGCC 45753
 QY 4102 CGGAGGAGAGTGGGTGATGGGCGGTGATTTGCGCAGCC 4161
 Db 45752 TCAAGCGGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45693
 QY 4162 CACGCGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4221
 Db 45692 GCTGCGGAGTGGGTGATGGGCGGTGATTTGCGCAGCC 45633
 QY 4222 CTTGCGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4281
 Db 45632 GCGGCTTGTGCGCGGTGATGGGCGGTGATTTGCGCAGCC 45574
 QY 4282 AGAGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4341
 Db 45573 --AACCGGAGAGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45516
 QY 4342 GCGGCGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4401
 Db 45515 CTTGCGTGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44549
 QY 4402 TGTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4461
 Db 45458 GGTGACCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45399
 QY 4462 CGGTGACACCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4521
 Db 45398 CGGTGATGCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45339
 QY 4522 CTTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4581
 Db 45338 CGCGGTGACACCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45282
 QY 4582 CGCGGTGAGAGAGAGCTTGAACGTGAGTGGCGGTGATTTGCGCAGCC 4641
 Db 45281 TGTGTCGAGCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45222
 QY 4642 GTGCTGTGCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4701

Db 45221 TGGGAGAGGCTTCACTGCTGCGCGGTGATGGGCGGTGATTTGCGCAGCC 45162
 QY 4702 CGGACGAGTGTGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4761
 Db 45161 GGTTCAGACCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45102
 QY 4762 TGTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4821
 Db 45101 TGTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 45042
 QY 4822 CTTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4881
 Db 45041 GCACTGAGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44982
 QY 4882 TGTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 4941
 Db 44981 CCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44925
 QY 4942 GGGGCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5001
 Db 44924 AGGAGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44865
 QY 5002 GCTG-----CGGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5055
 Db 44864 GGTGCGTGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44805
 QY 5056 CGGACCGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5114
 Db 44804 TGTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44745
 QY 5115 AGGAGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5174
 Db 44744 GCACTGAGTGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44685
 QY 5175 CGGACGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5231
 Db 44684 CGGACCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44625
 QY 5232 CTTGCGGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5291
 Db 44624 AGGAGCGGTGATGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44565
 QY 5292 CACTGCTGTGAGTGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5351
 Db 44564 GCACTGCGGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44505
 QY 5352 AGTGTGAGTGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5411
 Db 44504 CCACTGAGTGTGAGCGGTGATGGGCGGTGATTTGCGCAGCC 44445
 QY 5412 GCACTGCGGAGCGGTGATGGGCGGTGATTTGCGCAGCC 5454
 Db 44444 GGTGCGGCGGTGATGGGCGGTGATTTGCGCAGCC 44402

RESULT 11
 US-11-075-185-2/c
 ; Sequence 2, Application US/11075185
 ; Publication No. US20050266434A1
 ; GENERAL INFORMATION:
 ; APPLICANT: REEVES, CHRISTOPHER D
 ; APPLICANT: JULIEN, BRYAN
 ; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
 ; FILE REFERENCE: 010099_03
 ; CURRENT APPLICATION NUMBER: US/11/075_185
 ; PRIOR FILING DATE: 2005-03-07
 ; PRIOR APPLICATION NUMBER: US 60/551,103
 ; PRIOR FILING DATE: 2004-03-08
 ; PRIOR APPLICATION NUMBER: US 60/568,290
 ; PRIOR FILING DATE: 2004-05-04
 ; NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.3
SEQ ID NO 2
LENGTH: 14172
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-11-075-185-2

Query Match 1.0%; Score 89.6; DB 14; Length 14172;
Best Local Similarity 42.6%; Pred. No. 6.1e-09;
Matches 641; Conservative 0; Mismatches 859; Indels 6; Gaps 3;

QY 4167 TCATGCGCGGTGGCGGCGGTCCACCAAGATGATCACTGATGTTGCTCTCG 4226
DB TAGAGGCTCTCGCGAAAGCGCGCTCTCGATCGCGGTCTCAAGTTCGCGTTGTGCGC 8121
QY 4227 TCGCGCGGAGCGAGTGAACGCGCGGTACGTCGCGCACTGCGGTGTGTAAGC 4286
DB GCGATGAGCCGACGTCACGCGGACCTCTCGTCGCGCGCGACGCGGCGGACGACGCGC 8061
QY 4287 GCGGTGTGAACCGGTTGTCTCCCTGCTCGCGAAGCGCGTCAAGGAGTGAAGCC 4346
DB TCCGAGAGGCGCGCGAAGAGCTTGGCTGAGCGCGACCGCACTGCGGATCTCGTCG 8001
QY 4347 ATGCGCGCGCGCACTGCTCATCTTGGCGGTGCGGATCTCGGTGAACCACTTGTCC 4406
DB AGGAACAGGCTGCGCGGTGGCGCGCGCAAGCGCGCTCTCGCGCCCTTGGCGGTG 7941
QY 4407 GCGCGGATGCGGAAGTTGTGATGCGCGGATCGCGTTCGCGACAGCGCGTCTCGT 4466
DB GTGAAGCGCGCGCGCGCGCGCAACAGCTCGCTCG--AGCAAGCGGTTCGCGGACC 7884
QY 4467 ACCACCGCGCGCGCTCGAAGCGCGGTGACCGCTTGTGAGCGTGAACCACTCG 4526
DB GCGCGCGAGTTCACTCGGACGAGCGCGCTCGCGCGCGCGCGCTGCGTGTGACGCGCG 7824
QY 4527 GCGTCACCGGATCGCGCGCAAGCGGTCAACCGCTCGTCAGCGCGCGCGCGCGCG 4586
DB CGCGCGAGAGCTCTTGCCTCGGTGCGCGCTCGCGCGGTGATGAGGATCGATCTCGGAG 7764
QY 4587 TCGAAGAGAGCTTGAACCTGATGCGCGCGGAGATCTTGGCG--AGCGCTCAAGAGTC 4645
DB TCGCGAGCGCGGTTCAGAGATCGAAGAGCGCGCTTCACTGCGCGCGCGCGCGAGTC 7704
QY 4646 TGTTCGCGCGCGCAAGGTGACGCGCGCATCGCGCGGTCTGCGCGGTGAAGAGCG--CG 4703
DB TGTCTGTAAAGGTGGAGTCTGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7644
QY 4704 GCGAGGTGTCCGGGTGACCAAGCGGTGCGCGGTGACGTCGCGAGAACCGGTGTG 4763
DB GCGCGGTGCGCAAGGCGCGCTCGACGGTGAGCGGAGCGGTGAGCTCGAAGCGGCTTG 7584
QY 4764 AGTCGAGCGCGCTCGCGCGGTGCGCGGTGCGCGAGGATCTTGAAGCGGATGATCACT 4823
DB GTCAAGAGGTGTAGGCGCGCGCGCGCGAGATCGCGGTGCGAGGCTCCCGAAC 7524
QY 4824 TCACCGGTGAGGTCAACGCGCGCGCGACGACAGTTCCAGGCGCGCGGTGCGAGGTG 4883
DB GCGGTGATGAGATGACGCGGAGATCGGCGCGGTCTTGGGAGATCGCTCCAGAGCTCG 7464
QY 4884 GCGATGAGTGCCTGACCGCGCAAGGTGCGCGAGCGCGGCGCTCGAATCTCCCGCAAG 4943
DB AGCGCGCTCATGCGCGCGCGATGATGTCGAGTGAAGAGATGCAAGTCTCCGCGCG 7404
QY 4944 GCGCGCGCGGTGTGAGCGCATGTTGTTTCAGAGGCGCGCATCGAGCGCGCGCAAGCGC 5003
DB AGGAGCTGAGCGCTCGCGCGCGCGAGCTGCGCGCAAGCGCGCTGTAACCGCGCTTCTG 7344
QY 5004 TCGCGGTGCGCGATGTCGCGCGCGCGCGCATGTCAGGCGGTGAGGAAAGCGCTCGGAGCG 5063
DB AGCCGCGGTGAGAGAGCTCGCAAGCGCGCGCTCATGTGAGCATGAGAGCGGAGCG 7284
QY 5064 CCGAAGATCGCGAGTGGTGTGATCGCGCTTCAAGCGCGCTGCGCGGTGTGAGCGCGG 5123

DB 7283 CTATGAGCTTCTCTCGCGCGAGTACATGTTGAATGAGCTCCCGAGCGCGCGCTGT 7224
QY 5124 ACCGCGAGCGAGAGGCTGCGAGCTGATGTTGAAGTGAAGTGTGCGCGAGCACT 5183
DB CGAGCTGATTCAGCGCGCGGTGCTCTTGAATGATGCGGTGAGGAGCGAGCGAGCGC 7164
QY 5184 CGGTGATTCGAGCGAGTATCAACGCGGTGCGGAGCGCGCGCTGCTCGAAGTCCG 5243
DB CGGTGCGCTGCGCGAGCGCGCTTGTGCTGTAAGAACGCGCTGAGAGCTGTCTGAGAGCGC 7104
QY 5244 CGGCGACCTGAGAGAGATGATCCGCTTCTGTTGCGGTGAAGACCGCGCAACCTCTCG 5303
DB CGGCAATCATGCTCTGCTTCTCTGCGCGCACCGAGCGCGATGATGCTGCTCGCGC 7044
QY 5304 AGTCGAGCGGTGATGCGAGTCTGATGACGCGGAGCGTCAAGCAAGTATGAGTGAAG 5363
DB CTCTGATGATGCGCGGTGAGTGGCGCGGAGCGCGCTGAGACCGGACCGCGATGTCGCGC 6984
QY 5364 TCGGAGCGGTCTCTCGGAGCACTGCGGTGATGTTGTCGCGTGAACATGCACTGCGCG 5423
DB CGGACGCGCATGCGCTGATGCGGTGACCAACAGTTCGTGAGCGCGCTGAGAGTGT 6924
QY 5424 CCAACTGCGGAGCGTGTGCGCAACCGCGCTCGGTGCGAGTGAAGCGCGGTGAGG 5483
DB CGCGGTGCGGTGAGCAAGCGGATCGCGGTGCTCTCTCGACATGTTGACACCT 6864
QY 5484 TGCCTGATCTCTTGAACAGAGCGGAGCATGCGGTGTTGCGCGGAGAGAGCG 5543
DB GCTTGTGCGCGAGCGCTTGAAGCATGTTGAACGTCGCGCGGATGAAGGAGCGAGCT 6804
QY 5544 GTTGATTCAGAGACTCACTCGCGTGACTGTGTCAGAGAGCGCGAAGTCTGA 5603
DB CCGGAGCGCAACCGGCGGAGTGGGCTGTCGCGCTGCGCGTCAAGCGAGCGGAG 6744
QY 5604 AGTACTTCTTGTCTCTGTTTGAATGCGGTGTCGCGCGGATCAACCGCTGCGGTGA 5663
DB TGTTCACCTGCTGCTCCGCGCTGTTGGCGTTCACAGCGCGAGCGCGCGCGCGC 6684
QY 5664 TGTGCG 5669
DB 6683 CTTGCG 6678

RESULT 12
US-10-858-730-38/c
Sequence 38, Application US/10858730
Publication No. US20050255568A1
GENERAL INFORMATION:
APPLICANT: Bailey, Richard B.
APPLICANT: Blomquist, Paul
APPLICANT: Doten, Reed
APPLICANT: Driggers, Edward M.
APPLICANT: Madden, Kevin T.
APPLICANT: O'Leary, Jessica
APPLICANT: O'Toole, George
APPLICANT: Trueheart, Joshua
APPLICANT: Walbridge, Michael J.
APPLICANT: Yorgy, Peter S.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
FILE REFERENCE: 14184-030001
CURRENT APPLICATION NUMBER: US/10/858, 730
CURRENT FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/475, 000
PRIOR FILING DATE: 2003-05-30
PRIOR APPLICATION NUMBER: US 60/551, 860
PRIOR FILING DATE: 2004-03-10
NUMBER OF SEQ ID NOS: 364
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 2736
TYPE: DNA
ORGANISM: Streptomyces coelicolor

US-10-858-730-38

Query Match 1.0%; Score 89.2; DB 8; Length 2736;
Best Local Similarity 41.8%; Pred. No. 8.2e-09;
Matches 836; Conservative 0; Mismatches 1143; Indels 21; Gaps 4;

QY 3668 GAGGCGCGGTATGAGGCGGTGAGTTCGAGGCGGACCGGCTGTGCTTGGCGGTGCG 3727
Db 2135 GAGGCGCGGAGTATGAGTTCGAGGCGGACCGGCTGTGCTTGGCGGTGCG 2076
QY 3728 GATGCGGACCGGTGAGGCGGTGATCTTGGCGGTGAGGCGGCGGACCGGCGGCG 3787
Db 2075 GCGGCGGTGAGGAGGAGTTCATGCGGCGGTCCAGGCGGCGGCGGCGGCGGCGG 2016
QY 3788 TCGGCGGTGAGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3847
Db 2015 GCGGCGGCGGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1956
QY 3848 GTCCTTATGACGTCGACGACCTTCTGCGGTGTCGTCGAGGATCAGAACCGTTGCG 3907
Db 1955 TCCGCGGCGAGGCGGCGGAGTATGAGTATCTTTCGAGATGACCTGCGCTTCTGATACC 1896
QY 3908 TTTCGCTTATCCCATCTCGCGTTGATCAACGACGCGTGTGTGCGAGTCGATGAC 3967
Db 1895 TTGA---TCTGCGCTTCAAGGCGTCCAGGCGTGGCGAGGATGCGTGTGGG 1840
QY 3968 GTTGCACACCGGCGGAGTTCCTCACTGACACCGAGGCGCGGTGCGGAGGCGGAGCAC 4027
Db 1839 GCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1780
QY 4028 CCGTTCGCGGAGCACTCGGTCTGTGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4087
Db 1779 GCGGTGCGGCGACGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1720
QY 4088 CATCTGTGTGACGCGGCGGAGAGTATGAGGCTGTGCGACGACCTTCTCGCGCGAGAT 4147
Db 1719 GATACGCGCGAATTGAGAGAGTTCGAGTATGCGAGATGACCTCTGAGCGTCCGCGG 1660
QY 4148 CGCTTCGACGCGGTCACTGTCGATGCGGTGTGCGGTGTGCGGTTCACAGATGATCACTA 4207
Db 1659 CAGCGGCGACGAGCGCGGTATGAGAGGCGGTGCGGAGGAGTCTTCCAGATGAT----- 1606
QY 4208 CTGTGATGTTCTCTCTGTCGCGGCGGAGCGAGTGCAGGTCAGCGCGGCGGTACGTCGCG 4267
Db 1605 GTGCGGCGGCTTATGCTCTGTCGTCGTCCTCAGACGCGGACGATGCTTTCGCCCA 1546
QY 4268 CAGCTGTGTGTATGAGCGGTGTGACCGGTTGTCTCTGTCGTCGCGGACG 4327
Db 1545 GCGGCGGTGAGGTTCATGACCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1486
QY 4328 GTTGCAGGAGTATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4387
Db 1485 GTTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1426
QY 4388 CTGCGTATGACCACTTGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4447
Db 1425 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1366
QY 4448 CAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4507
Db 1365 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1306
QY 4508 GTTGCAGGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 4567
Db 1305 GTTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1246
QY 4568 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4627
Db 1245 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195
QY 4628 CAGCGGCTTCAAGT 4687
Db 1194 GAGCTGGAAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1135

QY 4688 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4747
Db 1134 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1075
QY 4748 GCAGAACACCGGTGTATGAGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4807
Db 1074 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
QY 4808 TGAAGCATGATCACTTCACTGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 4867
Db 1014 GTTGTCTTCAGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 955
QY 4868 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4927
Db 954 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 895
QY 4928 GAACTCCCGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4987
Db 894 GAGCAGTTTCTCGGTGCGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 835
QY 4988 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5047
Db 834 GTGATCATCTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 775
QY 5048 GAACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5107
Db 774 GTTCACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717
QY 5108 CCGTGTATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5167
Db 716 AGGTATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 657
QY 5168 CTGTGCGGAGCAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5227
Db 656 AGGTCTTTCAGGAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
QY 5228 CCGTGTGAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5287
Db 596 TTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 537
QY 5288 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5347
Db 536 AGGTATGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
QY 5348 AGGTATGAGTATGAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5407
Db 476 GTTTCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417
QY 5408 CAGTGCACCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5467
Db 416 GTTGCAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
QY 5468 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5527
Db 356 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
QY 5528 GCGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5587
Db 296 CCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
QY 5588 AGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5647
Db 236 TATGTCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177
QY 5648 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5667
Db 176 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 157

RESULT 13
US-11-205-109-1
; Sequence 1, Application US/11205109

```
Publication No. US20050287641A1
GENERAL INFORMATION:
APPLICANT: Farnet, Chris
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Staiff, Alfredo
TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
FILE REFERENCE: 3002-2US
CURRENT APPLICATION NUMBER: US/11/205,109
PRIOR FILING DATE: 2005-08-17
PRIOR APPLICATION NUMBER: US/09/976,059
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: US 60/239,924
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 88421
TYPE: DNA
ORGANISM: Actinoplanes sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2077)..(3078)
OTHER INFORMATION: ORF 1; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3118)..(4032)
OTHER INFORMATION: ORF 2; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4038)..(5048)
OTHER INFORMATION: ORF 3; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (6665)..(5814)
OTHER INFORMATION: ORF 4; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (7703)..(6693)
OTHER INFORMATION: ORF 5; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9464)..(8130)
OTHER INFORMATION: ORF 6; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9691)..(10761)
OTHER INFORMATION: ORF 7; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 8; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
FEATURE:
NAME/KEY: misc_feature
```

```
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 14; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70099)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (81809)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87372)..(86803)
```

OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc_feature
LOCATION: (87454)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

Query Match 1.0%; Score 87; DB 14; Length 88421;
Best Local Similarity 44.3%; Pred. No. 1.8e-08;
Matches 513; Conservative 0; Mismatches 625; Indels 21; Gaps 3;

```
QY 1083 CGAGAGAGCCGTTCTGCTCACTGGCTGGCGGTGATGCGCGACAGGACCCCGGTCTGACA 1142
DB 61180 CGAGAGAGCCGTTCTGCTCACTGGCTGGCGGTGATGCGCGACAGGACCCCGGTCTGACA 61239
QY 1143 GAGCCAGTACGGCGGTGCGGACATCTTCCGCCACAGTGAAGTACGAGGCTTCCGCCGA 1202
DB 61240 CGAGCGTACGGCGGTGCGGACATCTTCCGCCACAGTGAAGTACGAGGCTTCCGCCCGT 61299
QY 1203 CACCGCACCTTCTCTCCAGCCCAACCCGCTCATAGAGGGGGCCGACCCGACGCGCGG 1262
DB 61300 CCGCTTCCGCGCGCGCGCGGAGCGCATATGACCGCACCGGCGACCGGGTCCGGTGAACGC 61359
QY 1263 GATGATTCACGAGATGACCCGCGCGAGACCGGGCCCTGCGCAAGGTCTGACGACGCG 1322
DB 61360 CGACGCGACCTGCTGCTTCCGCGCGCGCGCGACGACGACGATGACGATCCGCGCTTCCG 61419
QY 1323 CTTCACCCCGCGTACGATCGCGCGACCTTCGACCGCGCGATCCGGAGGTGACCGCGTCT 1382
DB 61420 TATCGAGCCCGCGGAGGTGAGGCGCGTCTGCGCGCGCGCGACGCGCGCGCGCGCGGT 61479
QY 1383 GCTGCGCGACGCGCGGTGACCGCTTTCGACCTGCTGAGAGCGCGCTTCCCGCGCGCGT 1442
DB 61480 GGTGCTGCGCGCGGAGCGACCTTCCCGCGCGCGCGCGCGCGCGCTGCTGACGCTGTCGCG 61539
QY 1443 CACGATGCTGCGCGAGCTGCTGCGCGCTGCGCGCGAGTGAACACAGCATTTG---GTGA 1499
DB 61540 CGACCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGA 61599
QY 1500 CTGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1556
DB 61600 GCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61659
QY 1557 CGAAGCATCATGACAGGTGCTGAACCGCGCTCACTCTCTGACCTGCTGACAGGTGCTGCGA 1616
DB 61660 CAAAGCATCATGACAGGTGCTGAACCGCGCTCACTCTCTGACCTGCTGACAGGTGCTGCGA 61719
QY 1617 ACAGCGCGCGCGACCCCGCGGACGACCTGATCTCCCGCGTGTGCTGCGCGAGTCAACG 1676
DB 61720 CCGCGCGCGCGGTACCCACCGCGGAGAGGTGCTGCTGCGAGGTCTTCCCGCGAGGTGCTG 61779
QY 1677 GCGCACCTTCAACGAGTGAAGCGCGCACTTCTTCAACGAGCTGCTGCTGCTGCGGCGCA 1736
DB 61780 CCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 61839
QY 1737 CATCAACACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
DB 61840 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 61899
QY 1797 CTGAGCGCGCGCGCGCGGACCGCGGCTGATGCGCGCGCATGACGAGGAGTGTGCG 1856
DB 61900 CCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 61959
QY 1857 TTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1916
DB 61960 GCGCAACCTTATCCCGAGAGACCGCGCGCGCATACGCGCGCGCTGCTGCTGCTGCGA 62019
QY 1917 GGTGAGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC----- 1961
DB 62020 GCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 62079
QY 1962 GCGCAACCGCGATCCCGTGGCGCGATCCGACCGCGGACAGCTTGAACCGCGTCCCGCAAGT 2021
```

```
DB 62080 GCGCGCATCTATCCGCTCGCGCGCGCTCCAGAGGCGCATGCTTCCACGACCTATGCG 62139
QY 2022 CGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2081
DB 62140 CCGCGAGACGCGCGAGGACGCTTATCATGATGCGCGCGCGGTGTGAGTTGACCTCGCGGA 62199
QY 2082 GCGCGCGCTGAGAACGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2141
DB 62200 CCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62259
QY 2142 CTGACACCGCGAGACGACGAGTGTGCTGCTGACGAGATGCTGCTGCGACCGCGCA 2201
DB 62260 CACCGCGGTGCTGCGGACGCGCTGCGCGAGCGGTGAGGTGTGCGCGCGCGCGCG 62319
QY 2202 CTTCCCGGTGCTGCGCGCG 2220
DB 62320 CTTCCCGGTGACCGAGGTG 62338
```

RESULT 14

US-11-175-689-5
Sequence 5, Application US/11175689
Publication No. US20060024806A1
GENERAL INFORMATION:
APPLICANT: STINEAR, TIMOTHY P.
APPLICANT: COLE, STEWART T.
APPLICANT: LEADLAY, PETER F.
APPLICANT: SMALL, PAMELA L.C.
APPLICANT: JOHNSON, PAUL D.R.
APPLICANT: JENKIN, GRANT A.
APPLICANT: DAVIES, JOHN K.
APPLICANT: HAYDOCK, STEPHEN F.
TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING
TITLE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES
FILE REFERENCE: 03495.0329-01
CURRENT FILING DATE: 2005-07-07
PRIOR FILING DATE: 10/987,592
PRIOR APPLICATION NUMBER: 60/519,864
PRIOR FILING DATE: 2003-11-14
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 3.3
SEQ ID NO 5
LENGTH: 1314
TYPE: DNA
ORGANISM: Mycobacterium ulcerans
US-11-175-689-5

Query Match 1.0%; Score 86.6; DB 14; Length 1314;
Best Local Similarity 46.7%; Pred. No. 2.9e-08;
Matches 429; Conservative 0; Mismatches 469; Indels 21; Gaps 4;

```
QY 1252 CGAGCGCGGGAGTATCAACGAGATGACCCGCGGAGACACCGCGCGCGCGCGCGCGCG 1311
DB 316 CTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
QY 1312 GTGACGAGCGCTTCAACCCCGCGTACGATGCGCGACCTTGAACCGCGCGCATCCGAGGTG 1371
DB 376 GTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 435
QY 1372 ACCCGGTGCTGCTGCTG-----CGAGCGCGGTGACCGCGTTCGACCTGCTGAGGCG 1422
DB 436 GGTGCGCGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495
QY 1423 CTGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482
DB 496 TACTGCTTCCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555
QY 1483 CACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542
DB 556 CCAACCGGT---TCTCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
```


Query Match

Best Local Similarity 43.5%; Score 85.4; DB 14; Length 3240;

Matches 629; Conservative 0; Mismatches 811; Indels 7; Gaps 5;

7489 GCCGCGCGACCGGGAACCGTACCGCCAGTCGGGCGGAACTGGGTGGTGGTGAACACCGG 7548


```

Db 1954 CGGCGCCACCGTTAGCACACACACCGGTTGCCGACGCGCGCGTGCATTGCGG 1895
QY 8626 TCTGACCGCGGTGCGCGGTGTCGACGTGGAATTGTGAGTGGACACGCGCACGCGG 8685
Db 1894 AATTACGGTTGCGCGGTGTCGACGTGGAATTGTGAGTGGACACGCGCACGCGG 1835
QY 8686 CCAGGACATGCTGAGTACATGCGGACCTTCGACTGGGTGACCGAGACCAACACCAT 8745
Db 1834 CGCTGCGCGCGGTGTCGACGTGGAATTGTGAGTGGACACGCGCACGCGGCTT 1775
QY 8746 GTCTGGGACGACCTCTGCGCATGACACCACTTCACCGGACCTTCTTTCGCGCTGAT 8805
Db 1774 GTCCACCGGAGCGCGGAGACCGTGCACCGGTGCGGTGCGCGCGGTGCGCGC 1715
QY 8806 GAGCCCGGACTCGCTCATCGACCGGATGTCGAGTTCTGCGCTCTGCGGTCCGACT 8865
Db 1714 CGGACCGCGGTGTCGACGTGGAATTGTGAGTGGACACGCGCACGCGGCTT 1655
QY 8866 GATGCTGCGGAGCGCGTGAACCTTCGCGCGCGCGGATCGCGGCGCGGATCAACCGGACCC 8925
Db 1654 GCGCGCGCGCGCACCTTTCGCGGCGCGCGGTACCGGCGGCGCGCGGTGCGCG 1595
QY 8926 GCACGCGC 8932
Db 1594 CGGTGCGC 1588

```

Search completed: April 7, 2006, 03:50:34
 Job time : 3417.3 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:50 ; Search time 6263.93 Seconds
(without alignments)
11913.100 Million cell updates/sec

Title: US-10-611-442-1

Perfect score: 8942
Sequence: 1 ggcgcgttcgacccatccatg.....gtggagaccggagatcc 9024

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Published Applications NA Main:*

- 1: /cgn2_6/ptodaca/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodaca/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodaca/1/pubpna/US09_PUBCOMB.seq:*
- 4: /cgn2_6/ptodaca/1/pubpna/US10_PUBCOMB.seq:*
- 5: /cgn2_6/ptodaca/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodaca/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodaca/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodaca/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodaca/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodaca/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8942	100.0	9024	US-10-611-442-1	Sequence 1, Appl1
2	1309.4	14.6	12441	US-09-888-384B-3	Sequence 3, Appl1
3	1309.4	14.6	13613	US-09-861-289-3	Sequence 3, Appl1
4	1309.4	14.6	13613	US-09-860-846-3	Sequence 3, Appl1
5	1309.4	14.6	13613	US-09-836-821-3	Sequence 3, Appl1
6	1309.4	14.6	13613	US-10-271-889-46	Sequence 46, Appl1
7	1309.4	14.6	13613	US-10-398-605-3	Sequence 3, Appl1
8	1241.2	13.9	60136	US-10-205-032-1	Sequence 1, Appl1
9	1203.2	13.5	84428	US-10-229-148B-1	Sequence 1, Appl1
10	729.4	8.2	1565	US-09-793-708-24	Sequence 24, Appl1
11	729.4	8.2	1565	US-10-201-365-22	Sequence 22, Appl1
12	729.4	8.2	1565	US-10-160-539-24	Sequence 24, Appl1
13	729.4	8.2	1565	US-10-468-828-24	Sequence 24, Appl1
14	729.4	8.2	1565	US-10-468-828-24	Sequence 24, Appl1
15	722	8.1	1458	US-09-861-289-9	Sequence 9, Appl1
16	722	8.1	1458	US-09-860-846-9	Sequence 9, Appl1
17	722	8.1	1458	US-09-888-384B-9	Sequence 9, Appl1
18	722	8.1	1458	US-09-836-821-9	Sequence 9, Appl1
19	722	8.1	1458	US-10-271-889-52	Sequence 52, Appl1
20	720.6	8.1	85915	US-10-398-605-9	Sequence 9, Appl1
21	720.6	8.1	85915	US-10-647-196-1	Sequence 1, Appl1
22	667.2	7.5	1467	US-10-205-032-37	Sequence 37, Appl1
23	612.8	6.9	17596	US-10-611-442-2	Sequence 2, Appl1

C 24	596.6	6.7	1227	5	US-10-205-032-35	Sequence 35, Appl1
C 25	595.4	6.7	1248	3	US-09-861-289-7	Sequence 7, Appl1
C 26	595.4	6.7	1248	3	US-09-860-846-7	Sequence 7, Appl1
C 27	595.4	6.7	1248	3	US-09-888-384B-7	Sequence 7, Appl1
C 28	595.4	6.7	1248	3	US-09-836-821-7	Sequence 7, Appl1
C 29	595.4	6.7	1248	3	US-10-271-889-50	Sequence 50, Appl1
C 30	595.4	6.7	1248	3	US-10-398-605-7	Sequence 7, Appl1
C 31	519.2	5.8	5970	3	US-09-793-708-21	Sequence 21, Appl1
C 32	519.2	5.8	5970	6	US-10-201-365-11	Sequence 11, Appl1
C 33	519.2	5.8	5970	6	US-10-160-539-21	Sequence 21, Appl1
C 34	519.2	5.8	5970	8	US-10-468-828-21	Sequence 21, Appl1
C 35	519.2	5.8	5970	9	US-10-846-335-21	Sequence 21, Appl1
C 36	500.4	5.6	45055	6	US-10-107-431-277	Sequence 277, Appl1
C 37	496.6	5.6	1401	6	US-10-107-431-40	Sequence 40, Appl1
C 38	478	5.3	5975	9	US-10-796-304-1	Sequence 1, Appl1
C 39	474.2	5.3	1404	6	US-10-084-846A-103	Sequence 103, Appl1
C 40	474.2	5.3	59816	6	US-10-084-846A-2	Sequence 1, Appl1
C 41	474.2	5.3	59816	6	US-10-084-846A-2	Sequence 2, Appl1
C 42	463.4	5.2	30943	9	US-10-680-860A-1	Sequence 1, Appl1
C 43	461.2	5.2	1452	3	US-09-758-759-30	Sequence 30, Appl1
C 44	461.2	5.2	1452	10	US-11-021-825-30	Sequence 30, Appl1
C 45	461.2	5.2	1455	6	US-10-107-431-44	Sequence 44, Appl1

ALIGNMENTS

RESULT 1
US-10-611-442-1
Sequence 1, Application US/10611442
Publication No. US20040203015A1
GENERAL INFORMATION:
APPLICANT: Hutchinson, C. Richard
APPLICANT: Katz, Leonard
APPLICANT: Reid, Ralph
APPLICANT: Hu, Zhihao
APPLICANT: Gramajo, Hugo
TITLE OF INVENTION: RECOMBINANT GENES FOR POLYPEPTIDE
TITLE OF INVENTION: MODIFYING ENZYMES
FILE REFERENCE: 300622009100
CURRENT APPLICATION NUMBER: US/10/611,442
CURRENT FILING DATE: 2003-06-30
PRIOR APPLICATION NUMBER: US 60/393,016
PRIOR FILING DATE: 2002-06-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 9024
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)
OTHER INFORMATION: Genomic DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(9021)
OTHER INFORMATION: n = A,T,C, G, or none
US-10-611-442-1

Query Match 100.0%; Score 8942; DB 8; Length 9024;
Best Local Similarity 100.0%; Pred No. 0;
Matches 9024; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGTTCGATCCATGATCGCTTAATGTCGGTTCATTCCTTTGATGGGGAT 60
DB 1 GCGCGTTCGATCCATGATCGCTTAATGTCGGTTCATTCCTTTGATGGGGAT 60
QY 61 GATGTCATTCGAAATTCGAGACCGGTGTAAGCTTCGAGCTCGGCTGCTTTC 120
DB 61 GATGTCATTCGAAATTCGAGACCGGTGTAAGCTTCGAGCTCGGCTGCTTTC 120
QY 121 GTCAGTTCCTCCATCGCGCGAAGACCGGACGAGCGAAGAACGAAATGTTTCGGA 180

121 GTGAGTTCCCATGCGCGAAGAACGCGACGCGAACCGAAGAAATGTTCTCGCA 180
181 ATGCTCTCTGTCACAGCGCGACAGAGTGTTCGGGTCTCCGTCTCGTCTCGAGCTTGG 240
181 ATGCTCTCTGTCACAGCGCGACAGAGTGTTCGGGTCTCCGTCTCGTCTCGAGCTTGG 240
241 GACGACGAGCGCGCGTCAAGAGGCTCAACACCTTCAGACGTACATTCTCAATCTGCGC 300
241 GACGACGAGCGCGCGTCAAGAGGCTCAACACCTTCAGACGTACATTCTCAATCTGCGC 300
301 AAGATGTTCTGTCGCGGTCAACCGGTCTGCGCGCGAGGAGTCAACCGGAGTCTGCTCATC 360
301 AAGATGTTCTGTCGCGGTCAACCGGTCTGCGCGCGAGGAGTCAACCGGAGTCTGCTCATC 360
361 ACCCGGCGCGCGCGCTATCTGCTCTCGCGGTGACCGGATTCGCTCGACGTTAC 420
361 ACCCGGCGCGCGCGCTATCTGCTCTCGCGGTGACCGGATTCGCTCGACGTTAC 420
421 CAGGCTCTGATGTCGCGCGCGCTGCGCGCTCGGCTCGGCGACGAGTCAAGGACG 480
421 CAGGCTCTGATGTCGCGCGCGCTGCGCGCTCGGCTCGGCGACGAGTCAAGGACG 480
481 CGCAGACTCAACGAGCGCTCTGCGCTCTGCGCGCGCGCTCTGCTCGACGTTCCGCTG 540
481 CGCAGACTCAACGAGCGCTCTGCGCTCTGCGCGCGCGCTCTGCTCGACGTTCCGCTG 540
541 GGTGCGGTGCTGAGAGTGAAGCGTTCGCGAACCTGAGAGAGTCTGCTCATGAGCGAAG 600
541 GGTGCGGTGCTGAGAGTGAAGCGTTCGCGAACCTGAGAGAGTCTGCTCATGAGCGAAG 600
601 TACCTGTCGCGCGAGAGTTCGTCAGGAGAGTACCGCGAGAGGCTTCATGAGCTGAC 660
601 TACCTGTCGCGCGAGAGTTCGTCAGGAGAGTACCGCGAGAGGCTTCATGAGCTGAC 660
661 GCGCTCAACCGCGAGAGTTCGTCAGGAGAGTTCGCGAGAGGCTTCATGAGCTGAC 720
661 GCGCTCAACCGCGAGAGTTCGTCAGGAGAGTTCGCGAGAGGCTTCATGAGCTGAC 720
721 CATCTCAAGTGTGACCGCGCGCGAGCGTTCGAGAGTCTTCGACCGGTTGCGTTCGAC 780
721 CATCTCAAGTGTGACCGCGCGCGAGCGTTCGAGAGTCTTCGACCGGTTGCGTTCGAC 780
781 GTGACGAGAACTGAGGTCTGAGAACCGGAGCGCGAGGTGACAGAGTCAACGAGCGATCTG 840
781 GTGACGAGAACTGAGGTCTGAGAACCGGAGCGCGAGGTGACAGAGTCAACGAGCGATCTG 840
841 AACCGCGAGACGCTTCGAGAGAGTCTGCGCGTCACTCGTCTCGTTCGCTCGAGGTC 900
841 AACCGCGAGACGCTTCGAGAGAGTCTGCGCGTCACTCGTCTCGTTCGCTCGAGGTC 900
901 GCGCGCACGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 960
901 GCGCGCACGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 960
961 GACACCTCGAGACGCTTCGAGAGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1020
961 GACACCTCGAGACGCTTCGAGAGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1020
1021 CCGGTTGACGAGACGCTTCGAGAGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1080
1021 CCGGTTGACGAGACGCTTCGAGAGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1080
1081 GCGCGAGAGCGCTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1140
1081 GCGCGAGAGCGCTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAG 1140
1141 CAGAGACGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGT 1200
1141 CAGAGACGAGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGTTCGAGGT 1200
1201 GACACCGCGACCTTCGCTTCGAGACCGCGCTCATGAGGCGCGACCGCGACGCGC 1260
1201 GACACCGCGACCTTCGCTTCGAGACCGCGCTCATGAGGCGCGACCGCGACGCGC 1260

1201 GACACCGCGACCTTCGCTTCGAGACCGCGCTCATGAGGCGCGACCGCGACGCGC 1260
1261 GGGATGATTCACAGAGTTCGACCGCGCGAGACCGGCGCTTCGAGAGTTCGAGAGC 1320
1261 GGGATGATTCACAGAGTTCGACCGCGCGAGACCGGCGCTTCGAGAGTTCGAGAGC 1320
1321 GCGTTGACCGCGGTGACGATTCGCGACCTTCGAAACCGCGCATTCGAGAGTTCGAGAGC 1380
1321 GCGTTGACCGCGGTGACGATTCGCGACCTTCGAAACCGCGCATTCGAGAGTTCGAGAGC 1380
1381 GCGTTGACCGCGGTGACGATTCGCGACCTTCGAAACCGCGCATTCGAGAGTTCGAGAGC 1440
1381 GCGTTGACCGCGGTGACGATTCGCGACCTTCGAAACCGCGCATTCGAGAGTTCGAGAGC 1440
1441 GTGACGATTCGTCGCGCGAGTTCGAGGAGTTCGCGCGAGTTCGAGAGTTCGAGAGC 1500
1441 GTGACGATTCGTCGCGCGAGTTCGAGGAGTTCGCGCGAGTTCGAGAGTTCGAGAGC 1500
1501 TGGTTCGCGCGCGCTTCGAGCATTCAGATGAGACCGGACCGATTCGCGCGCTTCGAGAA 1560
1501 TGGTTCGCGCGCGCTTCGAGCATTCAGATGAGACCGGACCGATTCGCGCGCTTCGAGAA 1560
1561 GCGATTCATGAGAGTTCGAGACCGCGCTTCACCTTCGAGAGTTCGAGAGTTCGAGAGC 1620
1561 GCGATTCATGAGAGTTCGAGACCGCGCTTCACCTTCGAGAGTTCGAGAGTTCGAGAGC 1620
1621 CCGGCGCGACCGCGCGAGACCTGATTCGCGCGCTTCGAGAGTTCGAGAGTTCGAGAGC 1680
1621 CCGGCGCGACCGCGCGAGACCTGATTCGCGCGCTTCGAGAGTTCGAGAGTTCGAGAGC 1680
1681 ACCCTGACGAGAGTTCGAGAGGCGCGACCTTCGACAGCGTTCGCTTCGCGCGCGACATC 1740
1681 ACCCTGACGAGAGTTCGAGAGGCGCGACCTTCGACAGCGTTCGCTTCGCGCGCGACATC 1740
1741 ACCACACCGCTTCGCTTCGAGAGGCGCGACCTTCGACAGCGTTCGCTTCGCGCGCGACATC 1800
1741 ACCACACCGCTTCGCTTCGAGAGGCGCGACCTTCGACAGCGTTCGCTTCGCGCGCGACATC 1800
1801 ACCGCGCGCGCGCGAGACCGCGGTTCGATTCGCGCGAGTTCGAGAGTTCGAGAGC 1860
1801 ACCGCGCGCGCGCGAGACCGCGGTTCGATTCGCGCGAGTTCGAGAGTTCGAGAGC 1860
1861 CGCGCGCGCTTCGCGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 1920
1861 CGCGCGCGCTTCGCGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 1920
1921 GAGATTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 1980
1921 GAGATTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 1980
1981 GCGCATTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2040
1981 GCGCATTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2040
2041 TCCCTTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2100
2041 TCCCTTCGCGCGCGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2100
2101 GTGCGCTTCGAGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2160
2101 GTGCGCTTCGAGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2160
2161 ACGCTGAGTTCGAGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2220
2161 ACGCTGAGTTCGAGAGAGTTCGAGAGGCGCGACCGAGCGACCGCTTCGAGAGGTC 2220
2221 GTGACCGCGCGAGTTCGAGAGGCGCGACCGCTTCGAGAGGCGCGACCGCTTCGAGAGGTC 2280
2221 GTGACCGCGCGAGTTCGAGAGGCGCGACCGCTTCGAGAGGCGCGACCGCTTCGAGAGGTC 2280
2281 GTGACCGCGCGAGTTCGAGAGGCGCGACCGCTTCGAGAGGCGCGACCGCTTCGAGAGGTC 2340
2281 GTGACCGCGCGAGTTCGAGAGGCGCGACCGCTTCGAGAGGCGCGACCGCTTCGAGAGGTC 2340

[illegible]

QY	3421	GAACATGCACTCCGGGCTCCGGGGTGAACCCGACGCTGTAAAGGAAAGCCGGCTTCGCCCG	3480
Db	3421	GAACATGCACTCCGGGCTCCGGGGTGAACCCGACGCTGTAAAGGAAAGCCGGCTTCGCCCG	3480
QY	3481	TACCGACAGCTTCGAACGCGCCGGGGGGTTCAAGCGGCTCATGCTGTGTGCTCCAGTACTT	3540
Db	3481	TACCGACAGCTTCGAACGCGCCGGGGGGTTCAAGCGGCTCATGCTGTGTGCTTCAGTACTT	3540
QY	3541	GCCGCCGGGGCCCTGCTTCACCCGCCGTGACCAGCTCGGGAGTGCATGAAAGGTTCAG	3600
Db	3541	GCCGCCGGGGCCCTGCTTCACCCGCCGTGACCAGCTCGGGAGTGCATGAAAGGTTCAG	3600
QY	3601	CAGCTCCGCCAACCCCGGCTCCGGTCAAGTGAACCCAGCGGCATCTGTCTCAACGGGGT	3660
Db	3601	CAGCTCCGCCAACCCCGGCTCCGGTCAAGTGAACCCAGCGGCATCTGTCTCAACGGGGT	3660
QY	3661	GAACGGGAGGGCGCGTAAAGTGGGGGGTAAATGTGGCGAGCCGGAACCGCCGCTGTCCGCTCCGC	3720
Db	3661	GAACGGGAGGGCGCGTAAAGTGGGGGGTAAATGTGGCGAGCCGGAACCGCCGCTGTCCGCTCCGC	3720
QY	3721	GGTGTCCGATGTCGGCACCCGGTGAAGCGGGTGAACCTCGGCGCTGAGCGCCCGACACACGCG	3780
Db	3721	GGTGTCCGATGTCGGCACCCGGTGAAGCGGGTGAACCTCGGCGCTGAGCGCCCGACACACGCG	3780
QY	3781	CGGCGCGGTCCGGCGTCCGGGTGTGTCCGCTCCGCGATCTCCGTTCGTACCGCGGGTCCCGT	3840
Db	3781	CGGCGCGGTCCGGCGTCCGGGTGTGTCCGCTCCGCGATCTCCGTTCGTACCGCGGGTCCCGT	3840
QY	3841	AGGAGTGGTCTTTCATGCAACGTGTGGAACCTTTCGTGTGTCTGTGTGTGTGAGAGATCAAGAAC	3900
Db	3841	AGGAGTGGTCTTTCATGCAACGTGTGGAACCTTTCGTGTGTGTCTGTGTGTGTGAGAGATCAAGAAC	3900
QY	3901	CGTTGCGTTTCCGCTTGTCCCATCCCGCGTTTGAATCAACGACCGCTGTGTGTGTGCGAGTCC	3960
Db	3901	CGTTGCGTTTCCGCTTGTCCCATCCCGCGTTTGAATCAACGACCGCTGTGTGTGTGCGAGTCC	3960
QY	3961	GGATGACGTGCGACACCCCGGCGGATGTCTCACTGGAACACGAGGGGCGGTCGGAGGG	4020
Db	3961	GGATGACGTGCGACACCCCGGCGGATGTCTCACTGGAACACGAGGGGCGGTCGGAGGG	4020
QY	4021	CGAACACCCGTTCCGGCGAGCCCATCTCCGTCTGTGTCAAGCCGAGCGGTGCTCCGTGCGGT	4080
Db	4021	CGAACACCCGTTCCGGCGAGCCCATCTCCGTCTGTGTCAAGCCGAGCGGTGCTCCGTGCGGT	4080
QY	4081	AGGCGCTCATCTGGTGGCAGCGGGGGGAGAAATAGAGGCTGTGGAAGCAACCTTCCGCGC	4140
Db	4081	AGGCGCTCATCTGGTGGCAGCGGGGGGAGAAATAGAGGCTGTGGAAGCAACCTTCCGCGC	4140
QY	4141	GCAGGATCCGCTCGAGCCCGGTCAACGTGCATATGCCGGTGTGCGGTGCCGTCCACAGAGATGA	4200
Db	4141	GCAGGATCCGCTCGAGCCCGGTCAACGTGCATATGCCGGTGTGCGGTGCCGTCCACAGAGATGA	4200
QY	4201	TCACTGATCTGTAGTGTGCTCTCCTCGTCCGGCGGGAGCAGATGTCACGATGACGCCCGCTGA	4260
Db	4201	TCACTGATCTGTAGTGTGCTCTCCTCGTCCGGCGGGAGCAGATGTCACGATGACGCCCGCTGA	4260
QY	4261	CGTCCGCGACCTCGCTGTGTAAAGCGCGGTGTGAACCCGGTGTATCTCCCTGTGTTCGG	4320
Db	4261	CGTCCGCGACCTCGCTGTGTAAAGCGCGGTGTGAACCCGGTGTATCTCCCTGTGTTCGG	4320
QY	4321	CGAACCGGTGAGGGAGGTGAGCCCATATGGCCCGCGCGCATCTGCTCATCTTGTCCGTTGG	4380
Db	4321	CGAACCGGTGAGGGAGGTGAGCCCATATGGCCCGCGCGCATCTGCTCATCTTGTCCGTTGG	4380
QY	4381	TGCGGATCTCGGTGACCACTTGTCCGGGCGCATGTCGAAATTTGTGCATATGSCCCGATTC	4440
Db	4381	TGCGGATCTCGGTGACCACTTGTCCGGGCGCATGTCGAAATTTGTGCATATGSCCCGATTC	4440
QY	4441	GTTTCGCGCACAGGCGGTGTGCTGTATGCAACCGCCCGCGCTTCGAAAGGGGTGTACCGCTT	4500
Db	4441	GTTTCGCGCACAGGCGGTGTGCTGTATGCAACCGCCCGCGCTTCGAAAGGGGTGTACCGCTT	4500
QY	4501	TGTGTGCGTGAAGCTGAACCTTCGCGCTGACCGGATCCGCGCACCGGACGCTGCACCGG	4560

```
Db 4501 TGGTGGGTGGAAGCTGAACAACCTCGGCGTCAACGGATCCGCCACCGAGCTCAACCG 4560
Qy 4561 TCTGTGACGCCAGGCGGTGGCGGCGGTCAAGAAAGACTTGAACCTGTGTGTGGCGCGCA 4620
Db 4561 TCTGTGACGCCAGGCGGTGGCGGCGGTCAAGAAAGACTTGAACCTGTGTGTGGCGCGCA 4620
Qy 4621 TCTTTCGACGCGCTTCAACAGGTGCTGTGGCGGCCCAAGGTGTACGCCCAAGTCCGCG 4680
Db 4621 TCTTTCGACGCGCTTCAACAGGTGCTGTGGCGGCCCAAGGTGTACGCCCAAGTCCGCG 4680
Qy 4681 CGGTCTGCGGGGTGACAGAGCGCGGCAAGTGTGCGGGTCGACAGACCGGTGCGGGGT 4740
Db 4681 CGGTCTGCGGGGTGACAGAGCGCGGCAAGTGTGCGGGTCGACAGACCGGTGCGGGGT 4740
Qy 4741 CGAGCTGCGAGAAACAACGCTGTGTGATGTCAGACCAAGCTGCGCGGTGGCGGGCGCA 4800
Db 4741 CGAGCTGCGAGAAACAACGCTGTGTGATGTCAGACCAAGCTGCGCGGTGGCGGGCGCA 4800
Qy 4801 AGGTCAATTGACGGGATGATCACTTCAACGCTGACGTCACCGGCGCGGACAGCAAGTTCCA 4860
Db 4801 AGGTCAATTGACGGGATGATCACTTCAACGCTGACGTCACCGGCGCGGACAGCAAGTTCCA 4860
Qy 4861 GGGCGACGCTGCGCTTTCAGGCTGCGATGACAGTCCGCTTACCCGACAGTCCGCAACC 4920
Db 4861 GGGCGACGCTGCGCTTTCAGGCTGCGATGACAGTCCGCTTACCCGACAGTCCGCAACC 4920
Qy 4921 GGGCTCTGAATCTCCGACACAGGGGCCCGCTGTGTGTAGGCAAGTTGTTCAGGGGCC 4980
Db 4921 GGGCTCTGAATCTCCGACACAGGGGCCCGCTGTGTGTAGGCAAGTTGTTCAGGGGCC 4980
Qy 4981 ACTCAGGCGGGGACAGAAACGCTGCGCGGTGCGCATGTCGAGCGCGGCCACAGTGCAGG 5040
Db 4981 ACTCAGGCGGGGACAGAAACGCTGCGCGGTGCGCATGTCGAGCGCGGCCACAGTGCAGG 5040
Qy 5041 GGTGACGAGAACGCTTCGAGAACCGCGGAAGATCGCCAGATCGGTGCTGTCAGCGC 5100
Db 5041 GGTGACGAGAACGCTTCGAGAACCGCGGAAGATCGCCAGATCGGTGCTGTCAGCGC 5100
Qy 5101 CGTGCCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5160
Db 5101 CGTGCCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5160
Qy 5161 GTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5220
Db 5161 GTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5220
Qy 5221 GGCACCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5280
Db 5221 GGCACCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5280
Qy 5281 GTAGAACCGGCCACCTCTCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5340
Db 5281 GTAGAACCGGCCACCTCTCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5340
Qy 5341 GTCCAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400
Db 5341 GTCCAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400
Qy 5401 CGGCTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5460
Db 5401 CGGCTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5460
Qy 5461 CAGGTGTACAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5520
Db 5461 CAGGTGTACAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5520
Qy 5521 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5580
Db 5521 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5580
Qy 5581 CACGAGACGCGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5640
Db 5581 CACGAGACGCGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5640
Qy 5641 CCGGATTCACCCGCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5700
Db 5641 CCGGATTCACCCGCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5700
Qy 5701 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Db 5701 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Qy 5761 CGCCCCGACAGAAAGAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5820
Db 5761 CGCCCCGACAGAAAGAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5820
Qy 5821 TACCCGACGAGGCGCGGCTTCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5880
Db 5821 TACCCGACGAGGCGCGGCTTCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5880
Qy 5881 GTCCATGTTCACACGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5940
Db 5881 GTCCATGTTCACACGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5940
Qy 5941 CTTGCAATTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6000
Db 5941 CTTGCAATTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6000
Qy 6001 TTTGCGAGGAAACAGGAAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6060
Db 6001 TTTGCGAGGAAACAGGAAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6060
Qy 6061 GCGGTGTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6120
Db 6061 GCGGTGTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6120
Qy 6121 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6180
Db 6121 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6180
Qy 6181 GGCCTTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6240
Db 6181 GGCCTTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6240
Qy 6241 GAACCGGATCTCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6300
Db 6241 GAACCGGATCTCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6300
Qy 6301 GGTCCGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6360
Db 6301 GGTCCGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6360
Qy 6361 GTTCTCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6420
Db 6361 GTTCTCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6420
Qy 6421 CCGATGTCTGCGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6480
Db 6421 CCGATGTCTGCGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6480
Qy 6481 TCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6540
Db 6481 TCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6540
Qy 6541 GTTGACCGGATCCCGCATAGCTCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6600
Db 6541 GTTGACCGGATCCCGCATAGCTCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6600
Qy 6601 CCGGATGTGACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6660
Db 6601 CCGGATGTGACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6660
Qy 6661 AAGGTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6720
Db 6661 AAGGTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6720
```


Db 8881 GGTGACCTTGGCCGCCCCCGATCCGAGCCCGGGTCAACGGAAACCCCGACCCCGGATGCT 8940
|
Qy 8941 GTGGGGTCCGAGCTGCGCACCCCGGGCCCGGACAGCTTCTGTGATCTGTGGCCCA 9000
|
Db 8941 GTGGGGTCCGAGCTGCGCACCCCGGGCCCGGACAGCTTCTGTGATCTGTGGCCCA 9000
|
Qy 9001 GGAAGTGAAGACCGGAGGATCC 9024
|
Db 9001 GGAAGTGAAGACCGGAGGATCC 9024
|
RESULT 2
US-09-988-384B-3
Sequence 3, Application US/09988384B
Publication No. US20030073824A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,536US1
CURRENT APPLICATION NUMBER: US/09/988,384B
PRIORITY FILING DATE: 2001-11-19
PRIORITY APPLICATION NUMBER: PCT/US99/14398
PRIORITY FILING DATE: 1999-06-25
PRIORITY APPLICATION NUMBER: US 09/105,537
PRIORITY FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 3
LENGTH: 12441
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-988-384B-3
Query Match 14.6%; Score 1309.4; DB 3; Length 12441;
Best Local Similarity 67.7%; Pred. No. 8.4e-284;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;
Qy 2369 GGGCGGGTCGACGCGCTTCAAGCGCGGATCAAGCGCCCGATGCTATCCACCCGTCG 2428
|
Db 9472 GGGGGGATCGGGGCGGGGTGCGGGGTCAAGCGCAAGAGCCGCGGCTCTCCACGCGTCC 9531
|
Qy 2429 GCGACGTCCCGTTCGAGTTGTTAGCGCGGGCGGTCAACGACTGTGCGAACCCTGCAAG 2488
|
Db 9532 GCGGGGTGCGCTCCAGCTGTTCAAGCGGGCGGTGACGACCTGATCAAGCCGTCATG 9591
|
Qy 2489 AAGAACTGTCCTCCCGCGCTGCGGATCGATGCTGCGCGCCGACGTCGACGAAGTCGTGACG 2548
|
Db 9592 AAGTACTGTGCGCTGCGACGCGCCGACCTGCGCGCGCGCTGACGAAGTCCCTGACG 9651
|
Qy 2549 ACCGAGTCAGAGCTCCGCTGCGGGGTGACCGGCGCGGACGTCGAGCGGGTGGCCCGCC 2608
|
Db 9652 ACCTCGAGGAGGAGGTGTCGGGGGTGACGCGCGCCGCGATGTACGGGTGCGCGCGTCC 9711
|
Qy 2609 AGCCCGGGGAAACCGGCTCTCCCGGTACAGTGAAGTCTTCGAGGAGGTCCATCTGACCC 2668
|
Db 9712 AGGTGCGGGAAAGCCGCTCTCGGGTACAGTGAAGTGTGCGGAGGATGCACTGACCC 9771
|
Qy 2669 GCGACCTGCGATGAGGCGGTGCGGGCGCATGTTTCGGCGCGGATCCGAAACGACTGCGCG 2728
|
Db 9772 GCGACCTGCGGTGCGCGGTGCGGGCGCATGTTGCGGGCTTGTATCCGACGAGTTCCGCG 9831
|
Qy 2729 TCGACACCCCGACGCGAGCTCTCAACCGGTAAACCAAGTGTGCTGCAATGCGCGGGTCC 2788
|
Db 9832 TCGGCCCCGGGTGCGGAGGCTGTTCAAGGGGTAGCGGTGATGATGTGAGTCCGGGGGTG 9891
|
Qy 2789 CGCTGCGGGCGGTATCGACGAACCGGGCGAGGCGCTCCCGAACTCGGTCCGTTCCGCC 2848
|
Db 9892 CGCTGCGGAGCCCGCTCTCTCGAAGGCGTTGAAGGCGCTCTGAGGTCCGCGCGCTCC 9951
|
Qy 2849 TCGACAACTGCGGTGCTCCGACCGGCTGTAGTCTCTGCGGACGGTGACGAAGTCGAGC 2908
|

Db 9952 TGGGAGACTTGGCGCTGTCACGGCGCTGTAGTCTCCGGAATGTTGACGAATCTCATC 10011
|
Qy 2909 GGGCGGTGCGGGCTGACCTGCTTGAAGTTCGGCCATGAAGTCGACCAAGTGTGATGAGCGG 2968
|
Db 10012 GTCTTCCCTGCGCGGGGTGTTGAAGTTCGGCCATGAAGTCGACCAAGTGTGATGAGCGG 10071
|
Qy 2969 TTGGCCCGCCCGGACAGATGATGTGTTGAGGCGGAGCCGACCGGTGCGCCGCTTCC 3028
|
Db 10072 GAGGCAACGGCCCGGAGCAAGATGTAGCGGAACCGAGGTTGATTCGCGCATCTGCGCTCG 10131
|
Qy 3029 GCGCGCATTCGCAAGAAAGTCCCGAGGTTCTTCTGACCCGTTGAAAGCGCCACGTTG 3088
|
Db 10132 GCGCGCAGCTGTGGAAGCGCGCGACGAGTTCTCGCGACGCGCGGAAAGCGCGCTTCTTG 10191
|
Qy 3089 CCGGTGTGCGCTGTATCTGCTGTTGTTGAGCCCGGACAGGGAGGTACGGAACCGCGCG 3148
|
Db 10192 CCGGTGTGCTGTGATCTCTGCTGTGAGGCGGTGAGCGAGGTGCGGATGCGCTGC 10251
|
Qy 3149 AGGTCCACAGCGCGGCTGCGCGCGCGAGCTGTTCCGTTGAGGCGGAAAGCGCTTGTG 3208
|
Db 10252 AGGCCCCAGAGCGCGGCTGCGCTGCGCTGCGAGGTGCGTCCGTTGAGCGGAAAGTTCGTG 10311
|
Qy 3209 TAGACGCTGAGGGGAAAGCCCGCGCGCGCGCTGCGACACCAAGCCCGGATACCGCGG 3268
|
Db 10312 TAGACGCTGAGGGCGCAAGCGCGGTGCGGTGCGCGCGCGCGCTCCGAGCGCGGG 10371
|
Qy 3269 TTGCTCAACGTTCCAGGCGCGCGGAGAGGTACATGCGCGGTGCTCCGACGCGACG 3328
|
Db 10372 TTGCTGAGCGGCTCAAGCGCGCGGAGAGGTACATGCGCGGTGCTCCGAGGTATC 10431
|
Qy 3329 TCGTCATGACCGAGGTGAGGATCTCGTTGCGCGGACCAACCGACTCGGCTGCTACTTG 3388
|
Db 10432 TCGTCATGACCGAGCGGACATGCGCTTCCGCGGTGAGGGCGAGCGGCTGTACCG 10491
|
Qy 3389 GCACTGTCAACCGCTACCGCAAGTGAACAGCGAACTGCAAGTCCGGCGCGGGTAAAGC 3448
|
Db 10492 GCGCGGCTCAACGAGACGAGAGTGCACAGCGAACTGCAAGTCCGGCGCGGGTAAAGG 10551
|
Qy 3449 CCGAGCGTGAACGGGAAAGCGGCTCCCGCGTACCGCAAGCTCCAGCGCGCGCGCG 3508
|
Db 10552 CCGAGCGTGAACGGGAAAGCGGCTTCCGCGAGCGCGCGGTGAAAGACCGCGCGCTGT 10611
|
Qy 3509 TTCAAGCGGCTGATCGTGTGCTCCAGTACTTGGCCCGCGGCTCTCCACCGCGGTG 3568
|
Db 10612 TCGAGCGGAGAGAGGCTGTTTCCAGTACGCGCGCGGGCGGCTGACCGCGGTG 10671
|
Qy 3569 CCGAGCTCGGAGTGCATGCAACAGGTGAGACGCTCGCGAAACCGGCGCGGTCCAGG 3628
|
Db 10672 CCGAGCTCCGGGACCTGCGCGGACAGGCGGAGGCGCGCGGAGGCGGTCCCGGTGACG 10731
|
Qy 3629 TCGAACCGAGCGGCGCATCTGCTCCAGCGGGTGAACGGCGAGGCGCGTATGAGCGGTG 3688
|
Db 10732 CCGAGCTGCTGCGCGGCTCTCTCCAGCGGGTGAAGGGGCTGTTGCGGTACGCAACG 10791
|
Qy 3689 AGTTCCGAGAGCGGACCGGCTGTGTTCCGCGGCTGTCGATGCGGACCGGATGAGCGG 3748
|
Db 10792 AGCCGAGAGAGTGAAGCGGCGGTGCTTCCGCTCTGCTGCGGCGCAAGAGCGCGCG 10851
|
Qy 3749 GTGACCTTCGCGCTGAGGCGCCCGACCAAGCGCGCGCGGTGCGGCTGCTGTCGCGG 3808
|
Db 10852 GCGAGGGGTGCGGCGGAGCGGCTGACCGCGCGCCCGAGATCGGCTCCGAGGTGCGCGAG 10911
|
Qy 3809 TCCGCGATCTCCGCTGCTGATCGGCGGTGCGGCTGAGAGTGTCTTCACTGAGTGAAC 3868
|
Db 10912 CGTTCCGCGCGGCGGTGCGGAGGAGCGGCGGCGGTGATCGGAGGCTCAATCGTGG 10971
|
Qy 3869 CTTCGCGCTGCTGAGTGCAGAGTCAAGAACGTTGCGTTTCCGCTGTCCACTCCGCG 3928
|
Db 10972 GCGTGAATGTTGAGGGGCGCGAGCGGAGGCGGCGGTCTCCGCGGTGCG-----GCGC 11026
|
Qy 3929 GTTGAATCAACGACCGCTGCTGCTGAGTTCGAGTGAAGTTCGACACCGCGCGGATTC 3988
|

Db 11027 GGTCAAGTTCGAGCCGCGGGTCCGCGACAGACGAGGTCCGCGACCCGCGCATGTC 11086
Qy 3989 CTCACCTGACACCGAGGGGCGGTCGGAGAGGCGAGACCCGTTCCGCGAGCCATCTGGT 4048
Db 11087 GTTCCTCCCGATGCGGTGCGGTGCGGACAGACGCGCGCGCGAGGCGTTGGT 11146
Qy 4049 CTGTGTCAGCCGCGACCGGTGCTCCGTCGAGGAGGTCATCTGTGAGGCGGAGGA 4108
Db 11147 GTCCGCGAGCGGGGCGGTCGGCTCCCGCGGTACGCTCCAGCTCGTGGCAGCCGCGCA 11206
Qy 4109 GAAGTGAAGGCTGTGCGACGACCTTCTCCGCGCGAGGATGCTGACCGCGGTCA 4168
Db 11207 GAAGTGAAGGCGGCTGTGCGACGCGCTTCCAGGACCTTCAAGACGAGGTCCGCTG 11266
Qy 4169 GATGCGGAGTGGCGGTGCGTCCAGAGATGATCACTTACGTGAAGTGGTCTCTCTGTC 4228
Db 11267 GATGCGGAGTGGCTCTGTGATCTGACGATCACTTACGTGATGTTGTTAGGCGCTG 11326
Qy 4229 GGGCGGAGCGAGTGCACGCGTGCACGCGCGTACGTCGCGAGCTCGCTGGTGTAGACGC 4288
Db 11327 GCGGTGTGTCGCGCGCGACGAGACGCGGGAGAGTCCGCGAGTCTCGCGTGAAGCGGC 11386
Qy 4289 GTGTGTTACCCGCTTGTCTCCGTCGTCGCGAACGCGTCAAGGAGGTGAAGCCCAT 4348
Db 11387 GTGTGTCGCGCGGTCCGCTGATGATCTCGGAGAACGCTGAGGAGGATGAAGCCCAT 11446
Qy 4349 GGGCGGCGCGCATCTCGCTCATCTTTCGCTGTGTCGCGATCTCGGTGACCACTTGTCCG 4408
Db 11447 GGGCGGCGCGCTCTCGCTCATCTTTCGCTGTGTCGCGCGCGCGGTGTCGCGCGCGAG 11506
Qy 4409 GCGCATGCGCAAGTGTGTCATGAGCGCGCGCATCTTTCGCGCGAGCGCGCTCGTCGAGC 4468
Db 11507 GTGGAAGCGCAAGTGTGAGAGGCGCGCATCTTTCGCGCGAGCGCGCTCGTCGAGC 11566
Qy 4469 CACGCGCGCGCTCTGAAAGCGGTGACCGCTTGTGTCGTCGAGAGCTGAAACCTCGGC 4528
Db 11567 GAGCGGCGCGCTCTGAAAGCGGTGACCGCTTGTGTCGTCGAGAGCTGAAACCTCGGC 11626
Qy 4529 GTCAACCGGATCCCGCCACCGGACGTCACCGCTGTGTCAGCCGAGGCGGTGGCGCGCTC 4588
Db 11627 GTCCGCGAGGCTCCCGCGCGCGCGCGCTGACCGCGCACCGAGCGCGGTGGCGCGCTC 11686
Qy 4589 GAAGAAGAGCTTACCTGTGTGTCGCGCGCATCTTTCGCGCGCGCTCAAGAGTGTG 4648
Db 11687 GAAGTACAGCGCGAGCGCGCTGTGTCGCGCATCTTTCGCGCGCGCGCGCGAGG 11746
Qy 4649 TCGGCGCGCAAGGTGTACCGCGACGATTCGCGCGGTCTGCGAGGTGACGAGCGCGAC 4708
Db 11747 GCGGCGCGCAAGGTGTACCGCGACGAGCGCGAGGTGCGAGGTGACGCGCGCGCAC 11806
Qy 4709 GTGTGTCGCGGTTCACAGACCGGTTCGCGCGGTGTCAGTCCAGAACCGCGTGTAGTCC 4768
Db 11807 CTGTGTCGCGGTTCAGAGGTGTCGCGGTGTCCGCGGTGATGTCCGCAAGACCGCGGTGAGGCC 11866
Qy 4769 GAGCGAGCTTCGCGCGGTGTCGCGCGGTGTCGCGCGGTGTCATTTGAGGAGTATCACTTCAAC 4828
Db 11867 GATTCAGCGCATGTGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGT 11926
Qy 4829 GGTGACGTACCG 4888
Db 11927 GGTGAGGCG 11986
Qy 4889 GCAAGTCCGTTACCCGACGAGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4948
Db 11987 GGCATCCGAGACCCCG 12046
Qy 4949 GCGGTGTGAGGCGAGTGTGTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5008
Db 12047 GCGGTGTGAGGCGCATGCTGTGTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12106
Qy 5009 GTTCGCGATCTGTGCG 5068
Db 12107 GTTCGATGCGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12166

Qy 5069 GATCGCGATCTGTGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5099
Db 12167 GAATGCGAGGTTCGATTAAGCGCTTTTTCACG 12197

RESULT 3
US-09-861-289-3
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3

Query Match 14.6%; Score 1309.4; DB 3; Length 13613;
Best Local Similarity 67.7%; Pred. No. 8.4e-284;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

Qy 2369 GGGCGGATGACCGCGCTTCAAGCGCGCGATGAGCGCGCGATGATCCACCCGTCG 2428
Db 10159 GGGGAGATCGGGGCGGGTGGGGTTCAGCGAGAGACCGCGGCGCTCTCCAGCGCTCC 10218
Qy 2429 GCGACGTCCTCCGTTTCAGATTGTGTGAGCGCGCGCGGTCAACGATCTGTGAAACCTGAGG 2488
Db 10219 GCGGCGTTCGCTCCAGCTGTGTTCAGCGCGCGCGGTCAACGATCTGTGAAACCTGAGG 10278
Qy 2489 AAGAATCTGTCCTCCCGCGGTGCGGATGATGTCGCGCGCGCGCGCGCGCGCGCGCG 2548
Db 10279 AAGTACTCTGTCGCGGTGCGGATGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10338
Qy 2549 ACCGAGTGAAGCTTCGCGGTTCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2608
Db 10339 ACTTCGATGAGGAGGTGTCGCGGTGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10398
Qy 2609 AGCGCGGAGAAACCGCGCTTCGCGGTTCAGGTGACGTCCTCCGACGAGTTCGATTCGACC 2668
Db 10399 AGGTTCGAGGAAACCGCGCTTCGCGGTTCAGGTGACGTCCTCCGACGAGTTCGATTCGACC 10458
Qy 2669 GCACACTGAGATGAGGCGGTTCGCGGTGATCGTTTCGCGGTGATTCGCGCAACGCTGGCG 2728
Db 10459 GCACACTGAGGATGAGGCGGTTCGCGGTGATCGTTTCGCGGTGATTCGCGCAACGCTGGCG 10518
Qy 2729 TCGACACCCCGACGCGAGCTTTCACAGCGGTACCCAGGTCGATTCGATTCGCGCGGTTC 2788
Db 10519 TCGGCGCGGTTCGCGAGGCTTTCACAGGCGGTACCCAGGTCGATTCGATTCGCGCGGTTC 10578
Qy 2789 CGCTCGCGCGGTTCGACGAAACCGCGCGAGGCGCTTCGCGCACTCGGTTCGCTTCGCC 2848
Db 10579 CGCTTCGCGAGACCGCGCTTCGAAAGGCTTCGAGGCGCTTCGCGCACTCGGTTCGCTTCGCC 10638
Qy 2849 TCGACAACTGCGCGGTTCGCGAGCGCGGTTCGCGGTTCGCGCGGTTCGCGCGGTTCGCGAGC 2908
Db 10639 TCGGCGAGTTCGCTTCGTCAGCGCGCGGTTCGCGGTTCGCGCGGTTCGCGCGGTTCGCGAGC 10698
Qy 2909 GGCCTGTGCGGCTGACCTCGTTGATTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGAGC 2968
Db 10699 GTCTTCGCTTCGCGCGGTTCGTTGAGGTTCGCGGTTCGCGGTTCGCGGTTCGCGAGC 10758

OY	2665	TTGGCCCCGCCCGGACAGATGATGTGGTTGAAGCCCGAGCCGGACCCGGTCCCGCCCTTTG	3028
Db	10759	GAGGCACAGGCCCGGGAGACACAGATGTATAGCCAAAGCCGAGGTTATGCGACATCGCGCTCG	10818
OY	3029	GGCGGCATCCGCACAGAAAGTCCCGGAGGTTCTTCCGTACCCGTTCCGAAAGGCGCACCTTG	3088
Db	10819	GGCGGCGAGCTGTGTAAAGCCGGACGAGGTTCTTCGCGAGCGCGCGGAAAGCGGCTTCTTG	10878
OY	3089	CCGGTGGTGGCTTGGTACTGTCGTGTTTGAAGCCCGTACAGGAGGTTACGGACCGCGCGG	3148
Db	10879	CCGGTGGTGTGCTGTACTCTCTGTCGTGTTGAGGCGCGTATAGAGCGAGGGTGCGAATGGCGTG	10938
OY	3149	AGGTTCCACAGACCCCGGCTGGCCGCGACAGTCTGTTCGTTGATAGGCGGAAAGGTTGGTG	3208
Db	10939	AGGCCCCAGAGAGCGCGGCTGGCGCTCCAGAGGTGGCTCGGTGTAGCGCGGAAAGAGTTCTGTG	10998
OY	3209	TAGAAGGTCAGAGGAGAAAGCCCCCGCGGCGCGCGTGCAGACACAAACGCCCCGATACCCGG	3268
Db	10999	TAGAAGGTGGGCGCGCAGGCGCGTGGTCCGTGGCTGTGCGCGGCGCAGGCTTCCGAGGCGGG	11058
OY	3269	TTGGTCAACGGTTTCCAGGCGCGCCGAGAGGTTACATCGCCGTCGTGGGTTCTTCGACGGGACC	3328
Db	11059	TTGGTGAAGCGGCTCAAGCGCGCGCGAGAGATACATCGCCGAGGGGTTGCCCGCGGGATTC	11118
OY	3329	TGCTGTGATGACCCGAGGTGAGGATCTCGTTGCCCGGACACACCGACTCCGCGTGTACTTG	3388
Db	11119	TGCTGTGAAGACCGACCCGGAACATGGCGGTGGCCGCGTTCAGAGGGCGAGCCGGTGTGTAGCGG	11178
OY	3389	GCACCTGTACCCGTCACAGAGTGAACAGCGGAAATGCAACGTCGCGGCGGGGGTATAGAC	3448
Db	11179	GGCGCGGTACACCGGACCGAAGATGGCAGCGGAACTGCAAGTTCGGGCGGGGGTATAGG	11238
OY	3449	CCGACGCTGTACGGAAGGCGGGCTCCGCGCGTACCGGACGCTTCAACGCGCGCGGCGG	3508
Db	11239	CCGACGCTGTACGGAAGACGGGCTTCTCGGCGAGCGCGCGGTGTGAAGACGCGCGCGCTGT	11298
OY	3509	TTCAAGCGGCTGTGATCGTGTGTTGCTCCAGTACTTGGCCCGCGGCTTCTCTCAACGCGCGTG	3568
Db	11299	TGAGACGGGACACAGGTGTCTTCTTCAGTACCCCGCGCGGGCGCGGTCTCGAACCGCGCGTG	11358
OY	3569	CCGAGTCGCGGAGTCCGATTCGATCGAACAGGTCGAGCGAGCTCGCGAACCCGCGCGCGTCCAGG	3628
Db	11359	CGAGACTCCGGAACCTGCCCCCAACAGGGCGAGAGAGCGCGGAAAGCGTCCCGGTGACG	11418
OY	3629	TCGAACCGACCGGCGCATCTGCTCCACAGGGGTGAAACGCGGAGCGCGCGTGTAGTGGCGGTG	3688
Db	11419	CCGAGTGTGTGGCGGGCTCTCTTCACGCGGGGTGAAGGGGCGTTGGCGGTGACGACGGCG	11478
OY	3689	AGTTTCGCGAGCCGGAACCGGCTGTCTGTTCCGCGCGGTGTGATGTGCGACACCGGTATAGCGG	3748
Db	11479	AGCCCGGAAGAGGTGGCGGCGGTCGTTCCGCTCCGCGCTGTTCGCGGACGACAGAGGCGCGCGG	11538
OY	3749	GTGACCTCGGCGCTGTAGAGGCGCCGACACAGCGCGCGCGCGTGTGGGTCGAGGTCGTGCGCG	3808
Db	11539	GGAGAGGTGTGGCCCGACGACCGGTGACCGCCCGCCCCAGATGGGCTTCGAGGTGTGCGCAG	11598
OY	3809	TCCGCGATCTCCGTCGAGTACGCGCGGTCCGCGTAAAGAGTGTCTTCAATCGACGTGCGAAC	3868
Db	11599	CGTTTCGCGCGGGCGGTGGCGGAAAGGCGGGCGGTCATGTGGAGGTCATGTGG	11658
OY	3869	CTTTCGCGGTCGTGTGTGCGAGGATCAAGAACCGTTGCGTTTCCGCTTGTCCACTCCG	3928
Db	11659	GCGTGAATGTCTGTGGGGGCGCGAGCGGGGCGGGGCGGTGTCCGCGTGGC-----GCGC	11718
OY	3929	GTTCGATCAACGCAACCGCTGTGTGTGTGCGAGTTCGATTAACGTTCGACACCCCGCGGATGTC	3988
Db	11714	GGTCAAGTTTCGCGCGCGCGGTCGCGCAAGACGCAAGCGGTTCGCGACCCCGCGGAGATGTC	11778
OY	3989	CTCACCTGACACAGAGGGGCGGTCGAGAGGGCGAGCACCCGTTTCGAGAGCCACTCGGT	4048
Db	11774	GTGTGTGCGGATGGCGGTGCGGTCGCGACGGGACAGCACGCGCGCGCGAGGCTTGGGT	11838
OY	4049	CTGTGTCAACCCGACCGGTGGCTCCGTGCGGTATGGGCGTCACTGTGTGACAGCGGGGGA	4108

[illegible]

RESULT 4

OY	4169	GATCCCGGTGGCGGGTGGCGGTCCGTCACACAGAGAGATGACAGTACTGGTAGTGGTCTCTCCCTGGTC	4228
Db	119564	GATCCCGGTGGTGGCGGTCTGTCGTGATCTTGACAGATCAAGTACTGGTAGTGGTGTGTAAGGCCGTGG	120133
OY	4229	GGGCGGAGAGCGAGATGTCACCGGTGACCGCCGCGATCGTCGCGCAGCTCTCGTGGTGTAGACCGC	4288
Db	12014	GCGGTGCGTGGTCCGCGCAGAGAGACCGCCGGGAGAGGTCCGCGAGGTGCTCGCGGTAGAGCGGC	120737
OY	4289	GTGGTTGACCCCGGTGTGGTCTCCCTGGTCTCGCGGGAACGGTCCAGAGGAGGTGAGCCCAT	4348
Db	12074	GTGGTTGGCGCGGGTTCGCGGTGCAAGACCTCGGGAGAAAGCGCTCAGAGGAGGTGAGGCCCAT	121333
OY	4349	GGCGCGCGCGCAGCTCGCTCATCTTGACGGTGTGGTGGCCGACATCTCGGTGACCACTTGTCCGAG	4408
Db	12134	GGCGCGCGCGCGCTTCGCTCATCTTGGGTGTGGTCTCGCGCGCGCGCGCTGCGCCCGGCGAG	121939
OY	4409	GCCGATGCGCGAAGTTGTGATGACGCGCGCATCGCTTCGTCGCGCAGAGCGCGCTGTCGGTGAC	4468
Db	12194	GTCCAGCGCGAAGTTGTGAGAGGGCGCGGATCCCGGCGCGAGGTCCGCGCGTCCGTGGTGAAC	122555
OY	4469	CACCGCGCCCGCTTCGAAAGCGGTGACCCGCTTGTGTGGCTGTGAAGCTTGAAACCTTCGCGC	4528
Db	12254	GACCGCGCGCGCTTCGAAAGCGGTGACGCGCGCTTGTGTGGCTGTGAAGCTTGAAAGCTTCGCGC	123133
OY	4529	GTCAACCGGATCCCGGCGCAACCGGACGTCCACCCGTCGTGACACCCAGAGGCGGTGGCGCGCGTC	4588
Db	12314	GTCCCGCGAGGCTGCGCGCGCGCGCGCGCGCGTGCACCGCGCAGACCGAGGCGGTGGCGCGCGTC	123737
OY	4589	GAGAAGAGACTTGAACCTGTGTGGTCCGCGCGCATCTTTCGCGCAGCGCTTCACAGGTCTGG	4648
Db	12374	GAAATGACAGCGCGCAGCGCGCGTGTGTGTGGGAGACTTTCGCGAGCTGGTCCGCGCGCGAGGG	124333
OY	4649	TGCGGCCCAACAGGTGTATGCCGACGATCGCGCGGTCTCGCGGGGTGACAGCGCGCGAC	4708
Db	12434	GCGCGCCCAAGGTGTGACCGCGACGACGCGCGAGGTCCGGGGTGTGACCGCGCGCGCAC	124933
OY	4709	GTGGTCCGGGTGACCAAGACCGGTCCGCGGGTGCAGGTCCGACGAAACACCGGTGTGAGTCC	4768
Db	12494	CTGGTCCGGGTGAGGTTCGCGGTGTCCGGGTGCATGTCCGCGAAGACCGGGGTGAGGCC	125555
OY	4769	GAGCGAGCTCGCGCGGTGCGCGGTGGCGCGCGAGAGTCAATTGACCGGCATGTATCATCC	4828
Db	12554	GATCCAGCGCAGTGTGTGTGGGGGTGGGGGAGAACTGTATGAGCGGCATGTATCATCTTCGCC	126133
OY	4829	GGTGACGTGACCGGCGCGCAGACCAAGTTCAGAGGCGACGGTGGCGTTGACAGTGGCGAT	4888
Db	12614	GGTGAGCGCGCGCGGTGCGCGAGAGAGCTTGAGGCCCGCGCGTGGCTGTGACAGGTGGCAC	126733
OY	4889	GCAAGTCCGTTACCCCGAACAGGTGGGAGAACCCCGGCGCTCGAACTCCGCGACACAGGGGCC	4948
Db	12674	GGCAATGCCGAGACCCCGGAGACCCCGGCGACCGCTCTCTCGAATCTCGCGAAGAGCGGGCC	127333
OY	4949	GCGGTGTGAGACAGTGTGTGTTCAAGGCGCACTTCAGAGCGGCGCAGAGAACCGCTGCGC	5008
Db	12734	GCGGTGTGAGACAGCACTGGCTGTGAGGGGCCCGGTTCAGACCGCTCTCGTACAGCTGGCGCG	127933
OY	5009	GTCCCGCATCTGTCGCGCGGCGCCACGTCGACGGGAGGTGACAGAAACGCTTCGAGACCGCGGAA	5068
Db	12794	GTCCATGCGGTTGGGCGCGCCCAAGAGAGCGGCTGGTGCAGAAACCGCGCGGCGCGCGGAA	128533
OY	5069	GATCGCCAGATCTGTCGCGGTACGCGCTTCACG	5099
Db	12854	GATTCGAGGTCCGATTAAGCGCGCTTTTCACG	12884

```

? APPLICANT: Zhao, L.
? TITLE OF INVENTION: DNA encoding methymycin and pikromycin
? FILE REFERENCE: 600.438U51
? CURRENT APPLICATION NUMBER: US/09/836,821
? CURRENT FILING DATE: 2001-04-17
? PRIOR APPLICATION NUMBER: 09/105,537
? PRIOR FILING DATE: 1998-06-26
? NUMBER OF SEQ ID NOS: 43
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 3
? LENGTH: 13613
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
? US-09-836-821-3

```

Query Match	14.6%	Score 1309.4;	DB 3;	Length 13613;
Best Local Similarity	67.7%;	Pred. No. 8.4e-284;		
Matches 1850;	Conservative	0;	Mismatches 876;	Indels 5;
				Gaps 1;

OY	2269	GGGGGGGATCGAACCGCGCTTCAGACCGGCGGAGATACGAGCGCCCGATGCTGATCCACCGCTCG	2428
Db	10159	GGGGGGGATCGGGGCGGGGTGCGGGATCAGCGCCAGGAAGCCCGGGGCTCTCTCCAGCGCTTC	10218
OY	2429	GGGACGTCCCGCTTCGATTTGTTGAGCCCGGCGGCTCACCGACTGTGCGAAACCTTCGAGG	2488
Db	10219	GGGGCGTTCGGGCTTCAGCTGTTCAGGGCGGGCGGTGACGACCTGATCGAAGCGCTTCAGT	10278
OY	2489	AAGAACTTCGTCCCCCGGCTCGGATGAGTCTGCGGCGCGAGGTGACGAAAGTCTGTGACG	2548
Db	10279	AAGTACTCTGTCCGCTTCGACGGCGCGACCTCCCGCGCGCTTCGACGAAAGTCCCTGAGG	10338
OY	2549	ACCGAGTGCAGGGCTCCGGGTCCGGGGGTGACCCCGGCGGGGACGATGAGGGGTGCGCCCGCG	2608
Db	10339	ACCTCGGTGAGGGAGGGGTGCGGGGGTTCACGGCGCGCGCATGTGAGGGGTCCGCGCGCTTC	10398
OY	2609	AGCCCGGGGAAACCGGCGCTCCCGGTACAGGTAGACGTCTTCGAGCAGGTGATCTTGACCC	2668
Db	10399	AGGTTCGGGGAAACCGGCGCTCCCGGTACAGGTACAGTCCCGCGAGAGATCGACTTGACACC	10458
OY	2669	GCCACTTGCGGATGCGGCTCGGGCGGATCGTTTCGGCGCGGATCCGCGAACAGCTGGGCG	2728
Db	10459	GCGACTTGCGGATGCGGCTCGGGCGGATCGTTTCGGCGCGGATCCGCGAACAGCTGGGCGG	10518
OY	2729	TCGACACCCCGACGCGAGGCTCTCCACAGCGCTTACCGAGGTGCTCTGACATGCGCCGGGCGTC	2788
Db	10519	TCGGCGCCCGGTGCGCAGAGCTGTTCAGGGCGGTACCGGTAGTCGATGTGAGGTCCGCGGGGTG	10578
OY	2789	CGCTTCGGCGGATGATGACGAAACCGGGCGAGGCGCTCCCGGCAACTCCGATCCGTTGCGCC	2848
Db	10579	CGCTTCGGCGGATCCGCTCTTGAAAGCGGTGAGGGCTCTCTGAGAGCTTCGCGCCGCTCTCC	10638
OY	2849	TCGACCAACTGCGGCTGCTCCGACCGCTGTAGTCTTCGCGGACGGGTGACGAAGTGAAGC	2908
Db	10639	TGCGGCAAGTTCGCGGTGTCAACGGCGCGCTGTAGTCTTCGCGGAATGTTCAGAAAGTGAATC	10698
OY	2909	GGCGCGGTGCGGCTGGAAGCTGTTGAGTTCCGCGCATGATGCAACGAGTGCATGAGCGG	2968
Db	10699	GTCTTCGCTTACCCGCGGTGTTGAGGTCCGCGCATGATGCAACGAGTGCATGAGCGGCGG	10758
OY	2969	TTGGTCCCGGCCCCGGCAGAGATGATGTGTTGAGGCGCGACCGGACCGGTGCGCGGCTTG	3028
Db	10759	GAGGCAACCGCCCCGGGACACGATGTAGCGCAACCGAGTTGATCCGACACTGCGGCTCG	10818
OY	3029	GGCGGCAATCCGACGAAAGTCCCGGAGGTTCTTCCTGACCTGTTGAAAGCGCCACACTTG	3088
Db	10819	GGCGGCAAGTCTTGAAAGCGGCGCAGAGTTCTTCGCGGACCGGCGGAAAGCGGCGCTTCTTG	10878
OY	3089	CGGATGTGAGCTGTGATCTGTCTGTTTTCAGCCCGGTACAGGAGGTACGACCGGCGCG	3148
Db	10879	CGGATGTGAGCTGTGATCTGTCTGTTTTCAGGCGCGGTACAGGAGGTACGACCGGCGCG	10938
OY	3149	AGGTTCACAGGCGCGGCTGCGCGGACGCTCTGTTCCGTTGAGGCGAAAGGCTGTGGTG	3208

Db	10939	AGGCCCAAGAGCCGGGGCTGGCCGCTCAAGGGGTGGCGCTCGGTGAGACCGAAGAGTTCCGG	10992
QY	3209	TAGACGGTCAGAGGCGAAGCCCGCCCGGGCGCGTGGCAACCAACGCCCGCATACCCGGG	3268
Db	10999	TAGACGGGTGGGCGCCAGAGCCGCTGGCTCGGTGGCGGTGCGCGCACAGCTCCGAGAGCCGGGG	11058
QY	3289	TTGGATCAACGGTTTCCAGGGCCGCGCGGAGAAGTATATGCGCGCTCGGGTTCTCCACCGCACCC	3328
Db	11095	TTGGTAGCGGGCTCCAGGCGCGCGCGAAGATCATTCGCTCGAGGAGTGTGCCCGGGGTATTC	11118
QY	3329	TCGTGATGACCGAGAGTGAAGGATTCGTTCCGTCGGGCGACCAACCCACTCCGCGTGTACTTG	3388
Db	11119	TCGTGATGACCGAACCGGAACATATGCGCTTTCGCGCGCTCAGAGGCGGACGGGTGTAGCGG	11178
QY	3389	GCACCTGTCAACCCGTCACGCGAAGTGAACGCGGAATATGCAATGCAACTGTGGCGCGGGTAGAC	3448
Db	11179	GGCGCGGATCAACCGGACGGAAGATGGCAGCGGAACTATGCAAGGTTCGGGCGCGGGGTAGAG	11238
QY	3449	CCGACGCTGTATCCGGGAAGCGGGGTCCCCCGCGTACCGCAACCTTCAACGCGCGCGCGGG	3508
Db	11239	CCGACGCTGTATCCGGGAAGACGGGGTTCCTGAGCAGCGCGCGCGTGTGAAGACGCGCGCTGT	11298
QY	3509	TTCAACGGGCGTGAATCGTGTGTCTCAAGTACTTGCGCGCGCGGGCGCTGTCAACCGCGCGTG	3568
Db	11299	TCGAGCGGGAGACAGGGGTCTTCTTCAAGTATGCCCCCGCGGGCGCGGTCTTGACCGCGGTG	11358
QY	3569	CCGACGCTGCGGGATGCAATCGAACAGGTTCAGACACTCGCGAACCCCGCGCGCGTTCACG	3628
Db	11359	CCGACGCTCCGGGACCTGCCCCGAACAGGGCGAGGAGGCGCGGAAAGCGTCCCGGTGAGAG	11418
QY	3629	TCGAACCGACGCGCGCAATCTGTCTCAACGCGGGTGAACGCGAGCGCGCGTATGTAGGCGGTG	3688
Db	11419	CCGACGCTGTGTGGCGGGCTCTCTTCCAGCGGGGTGAAGGGGCGTGTTCGTTAGCGCACGCGG	11478
QY	3689	AGTTTCGGCGAGCGGACCGCGCGTGTCTGTTGGCGGGGTGTGGAATCGGCAACCGGTGAGCGG	3748
Db	11479	AGCGCGACGAGATGTGGCGGGCGCGCGTGTTCGCGCGCTCGTCCGAGCGACAGAGCCGCGGCG	11538
QY	3749	GTTGACTTCGGCGCTGAAGCGCGCCGCAACCAACGCGCGCGCGTGTGGGTCTGTGCGCGCG	3808
Db	11539	GCGAGGGGTCTGGCGCGGACCGGGGTGGAACGCGCGCCGCCAAGATCGGCTCCGGGGTGGCGGCG	11598
QY	3809	TCCGCGATCTTCGTCGCTGATGCGCGGGTTCGCGTGAAGTGTCTTCAATCGACGTGCGAACCC	3868
Db	11599	CGTTTCGGCGCGGGCGGGTGGCGGAAGAGGCGGGGCGGCTATCCGAGAGCGTCAATCTGTGG	11658
QY	3869	CTTTCGCGCTGTGTGTGTGCGAGGATCAACGAACGGTTGCGCTTTCGCTTCCACTCCGCG	3928
Db	11659	GCGGTGATGTCTGGGGGGGCGCGAGCGGGGCGGGGCGCGGTGTGCGGTGGCG----GCGC	11713
QY	3929	GTTTGATCAACGCAACCGCTGTGTGTGTGCGAGATCGATGATACGTGCGCAACCCCGCGGATGTC	3988
Db	11714	GGTCAATTCGCGCGCGCGGGGTCCGCGCAGAGAACCGACAGGTTCGGCGAACC CGCGGATGTC	11773
QY	3989	CTCACTGGAACCCGAGGGGCGCGGTGGGAGGGCGAGCAACCGTTTCGCGAGCAACTCCGT	4048
Db	11774	GTCGTCGCCCATGAGCGGTGCGGTGCGGTCGAGGACAGACACGCGCGCGCGAGGGTTCGCT	11833
QY	4049	CTGTGTATGACCGCAGCGGTGTCTTCGTGCGGTAGGCGCTCATTTGGTGTGGCAGCGGGAGAA	4108
Db	11834	GTTGGGACAGCGGGGCGTGGCGGCTGCGCCGCGGTACGGCTCCAGCTCTGTGGCAACCCGCGCA	11893
QY	4109	GAAGTATGAGCGCTGTGTGCGACGACTTCTTCGCGCGCGAGAGATTCGCTGACGCGCGATCGTTC	4168
Db	11894	GAAGTATGAGCGCGGGGTGTGACACGCGCTTTCGCGCTTCAAGAACCTTCATATACAGAGTTCGCGGTG	11953
QY	4169	GATGCGCGGTGGCGGTGCGCTGCTGATCTGACGATCACTACTGTGTGTGTTGTTGAGGCGGTG	12013
Db	11954	GATGCGCGGTGGCGGTGCGCTGCTGATCTGACGATCACTACTGTGTGTGTTGTTGAGGCGGTG	12073
QY	4229	GGGCGGGAACGCAATGACAGGTGATACGCGCGTATCGTGGCGAGTTCGCTGTGTGTAGAGCGC	4288
Db	12014	GCGGTGTGTGTGCGCGACGAGAACCGCGGGAGAGGTTCGCGAGAGTGTCTCGCGGTAGAGCGCGC	12073

QY	4289	GTGGTTAACC	CGGTGTGG	CTCC	CGGTCT	CGGGGA	ACGG	TCAGAGG	AGGTGA	AGCC	CA	4348														
Db	12074	GTGGTTGCGG	CGGTTC	CGGTG	ATGA	CTCGGG	AAACGG	CTCAGAGG	AGGTGA	AGCC	CA	12133														
QY	4349	GGCGCGG	CGCACT	CGCTCA	CTTTC	CGGTGG	ATGCG	CGCATCT	CGGTGA	CA	CTTGT	CCG 4408														
Db	12134	GGCGCGG	CGGCGCT	CGCTCA	CTTTC	CGGTGG	ATGCG	CGCATCT	CGGTGA	CA	CTTGT	CCG 1219														
QY	4409	GCCGATCC	CCGAAT	GTTTGG	CAATGG	CGCCGA	ATCC	CTTTCGG	CCAGAG	CGCTTC	TCGTGG	AC 4468														
Db	12194	GTCCAA	CGCAAG	TTGGAG	GGGGCG	CGAATCC	CGGGCG	CGAGG	TCGCG	CGTC	TCGTGG	AC 1225														
QY	4469	CACGCGCC	CGCCCT	CGAAG	CGGG	GAACCG	CGCTT	GGTGG	ATGCG	CTTGG	AAAGCT	GA	CA	CTT	CGG 4528											
Db	12254	GACGCGCC	CGCCCT	CGAAG	CGGG	GAAT	GAAG	CGCTT	GGTGG	ATGCG	CTTGG	AAAGCT	GA	CA	CTT	CGG 1231										
QY	4529	GTCA	CCGGAAT	CGCGCA	CCGGA	CGTCA	CCCGT	CTG	TGACAG	CCCA	GGGCGT	GGCGG	CGTC 4588													
Db	12314	GTCCCG	CAAGGCT	CGCGCG	CGGCGCG	CGCTGA	CCGCG	CAAGCC	CAAGCG	CGGCGG	CGTC	CGGCG	CGTC 1237													
QY	4589	GAAGA	AGGCT	GA	CTGG	ATGG	CGGGG	AGATCT	TTC	CGACGG	CGCTT	CA	CA	AGT	CTGG 4648											
Db	12374	GAAGT	TA	CA	CGCG	AGCG	CGGTG	CTG	TC	TGG	AG	CA	CTT	TC	CG	AG	CGTGG	CGCG	CA	GGG 12433						
QY	4649	TCGCGCC	CA	AGGT	TA	ACGCG	GA	CGAT	TCGCGCG	CGTCTG	CGGG	ATGA	CA	GA	CGCGCG	CGAC 4708										
Db	12434	CGCGCC	CA	AGGT	TA	ACGCG	GA	CGAT	TCGCGCG	CGTCTG	CGGG	ATGA	CA	GA	CGCGCG	CGAC 1249										
QY	4709	GTGGT	CCGGGT	CCGA	CA	GAC	CCGGT	CGCGGG	TGA	CGT	CG	AGAA	CA	CG	GT	TA	AGT	CC 4768								
Db	12494	CTGGT	CCGGGT	CCGA	CA	GAT	CGGT	CGCGGT	TGA	CGT	CG	AGAA	CA	CG	GT	TA	AGT	CC 1255								
QY	4769	GAGC	CAC	GT	CGCG	CGGTG	CGCGGG	AGG	CGG	AAAG	GT	CA	TT	GA	CG	GA	TA	TC	AT	CAC 4828						
Db	12554	GAT	CA	CA	CGCA	GT	TCG	CGGG	TGG	CGCG	AA	CT	CA	T	GA	CG	GA	TA	TC	AT	CAC 1261					
QY	4829	GGTGA	CGT	CA	CCG	CGCC	GA	CA	CA	ATT	TC	CA	GG	CGA	CG	GT	TC	CA	GG	TC	GA	AT 4888				
Db	12614	GGTGA	CGC	CGCG	CGGTG	CGCG	GA	CA	CA	ATT	TC	CA	GG	CGA	CG	GT	TC	CA	GG	TC	GA	AT 1267				
QY	4889	GCA	GT	CGCG	TA	ACCC	CA	CA	AG	TC	CGCG	GA	ACCC	CGGG	CT	CG	AA	TC	CC	CG	CA	CA	CGGG	CCC 4948		
Db	12674	GGA	TTG	CGCG	CA	ACCC	CGCG	GA	AG	CT	CGCT	CT	CG	AA	TC	CC	CG	CA	CA	CGGG	CCC 12733					
QY	4949	GCC	TT	GGT	AT	GAG	CA	GT	TT	GT	TT	CA	AGG	CG	CA	CT	CA	AG	CGGG	CG	CA	GA	CA	CGG	CTG	CGG 5008
Db	12734	GCC	TT	GGT	AT	GAG	CA	GT	TT	GT	TT	CA	AGG	CG	CA	CT	CA	AG	CGGG	CG	CA	GA	CA	CGG	CTG	CGG 1279
QY	5009	GT	CG	CC	CA	AT	CGT	CGCG	CG	CA	CT	TC	GA	GGG	AT	CG	AG	AA	CG	CT	CG	GA	CG	CG	CA 5068	
Db	12794	GT	CG	AT	CGG	TT	GG	CG	CG	CG	CA	CT	TC	GA	GGG	AT	CG	AG	AA	CG	CT	CG	GA	CG	CG	CA 1285
QY	5069	GAT	CG	CA	GA	AT	CGG	TT	GG	AT	CGG	CT	TC	CA	GG	CT	CG	GA	CG	CT	CG	GA	CG	CG	CA 5099	
Db	12854	GA	AT	CG	CA	GA	AT	CGG	TT	GG	AT	CGG	CT	TC	CA	GG	CT	CG	GA	CG	CT	CG	GA	CG	CG	CA 12884

```

RESULT 6
US-10-271-889-46
; Sequence 46, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600,582US1
; CURRENT APPLICATION NUMBER: US/10/271, 889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18

```

```

; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-10-271-889-46

Query Match      14.6%  Score 1309.4;  DB 6;  Length 13613;
Best Local Similarity 67.7%;  Pred. No. 8.4e-284;
Matches 1850;  Conservative 0;  Mismatches 876;  Indels 5;  Gaps 1;

Qy      2369  GGGCGGGTCGACGCGCTTCAGACGCGCGGATCAGGCCCGCCGATGCTGATCCACCCGCTCG
Db      10159  GGGGGGATCGGGGGGGGGTGGGGGTCAAGCGAGAAAGCGGGGCTCTCCAGCGCTCC
Qy      2429  GCCAGCTCCGCTTCGAGTTGGTTGAGCCGGGCGGTCACCGACTGCTGAAAACGTCGAG
Db      10219  GCGCGCTCGCGCTCCAGCTGGTTCAAGCGCGGGGTGACACTGATCGAAAGCCGTCATG
Qy      2489  AAGAAGCTGTCCTCCCGGCTCGGATCGATGCTGCGGCGCGACGTCGACGAAGTCGTGAGC
Db      10279  AAGAGTCTGTCGCGCTTCGACGCGCGCACTCGCGCGCGCTGACGAAAGTCCCTGACG
Qy      2549  ACCGAGTCAGAGCTCCGCTCGGGGTGACCCGCGCGGACGTCGAGCGGGTGGCCCGCC
Db      10339  ACCTCGGTCAGGAGGTGTGGGGGTCAAGCGCGCGCGGATGTCGGGTGCGCCGCTCC
Qy      2609  AGCCCGGGAAACCGGCTCCCGGTACAGGTACGTCCTCCGACAGTCGATCTGACAC
Db      10399  AGCTCGGGAAAGCGGCTCCCGGTACAGGTACGTCCTCCGACAGTCGATCTGACAC
Qy      2669  GCCACTGTCGAGTGGGCGGTGGGGGATCGTTCCGGCGGGATCCGCAACGCTGGGCG
Db      10459  GCGACTGTCGAGTGGGCGGTGGGGGATCGTTCCGGCGGGATCCGCAACGCTGGGCG
Qy      2729  TCGACACCCCGACGAGCTCTCCACGCGTCGACCCAGGTGCGTCTGTCATCCCGGGCTC
Db      10519  TCGGCCCCCGATGCGCAGGCTCTTCAGCGCGTCGACCGTCGATGTCGATGTCGAGTCG
Qy      2789  CGCTCGCGGCGTACGACCAACCGGCGAGGCGCTCCCGCAACTCGCTCCGTTCCGCC
Db      10579  CGCTCGCGGACCCGCTCTCCAGAGCGTTGAGGGGCTCTGAGGCTCGGCGCTCTCC
Qy      2849  TCGACAACTGCGGTCGTCGCGACCGCTGTCGTCGTCGCGGACGTCGACGAGTCGAGC
Db      10639  TCGGCGAGCTGCGGTCGTCGCGACCGCTGTCGTCGTCGCGAGTCGACGAGTCGATC
Qy      2909  GCGCGGTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      10699  GTCTCGGCTTCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      2969  TTGCGCGCGCGCGCGAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      10759  GAGGACAGGCGCGGAGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3029  GCGCGGATCCGCGAGATCCCGGAGTTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      10819  GCGCGGATCCGCGAGATCCCGGAGTTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3089  CCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      10879  CCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3149  AGGTCCTCAAGCGCGGTCGCGCGGACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      10939  AGGTCCTCAAGCGCGGTCGCGCGGACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG

```

```

Qy      3209  TAGACGTCAGAGCGAAGCCCGCGCGCGCTGCGACCAACGCCCGGATACCCGCG
Db      10999  TAGACGTCAGAGCGCGAGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3269  TTGCTCAACGCTTCGAGCGCGCGGAGAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11059  TTGCTCAACGCTTCGAGCGCGCGGAGAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3329  TCGTCATGACCGAGGTCGAGATCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11119  TCGTCATGACCGAGGTCGAGATCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3389  GCACTGTCACCCGTCACGAGAAAGTACAGCGGAAATGACGTCGCGCGCGGTAGAGC
Db      11179  GCGCGCGTCACAGGACGAGAGTGGACAGCGGAAATGACGTCGCGCGCGGTAGAGG
Qy      3449  CCGACGCTGTCAGGAAAGCGGCTCCCGCGTCACGAGCTTCACCGCGCGCGCGCG
Db      11239  CCGACGCTGTCAGGAAAGCGGCTTCGCGCGAGCGCGCTTCGAGAGCGCGCGCTGT
Qy      3509  TTGAGCGGCTGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11299  TCGAGCGGACGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3569  CCGAGCTGCGGATTCGATGAAACAGGTCGAGCAGCTCGCAACCCGCGCGGTCCAGG
Db      11359  CCGAGCTGCGGATTCGATGAAACAGGTCGAGCAGGTCGCGGAGCGGTCGCGGTCCAGG
Qy      3629  TCGAACGAGCGGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11419  CCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3689  AGTTTCGCGAGCGGACCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11479  AGCGGACGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3749  GTGACTCGGCGTCGAGCGCGCGCAACGCGCGCGCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11539  GCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3809  TCGCGATTCCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11599  GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3869  CTTTCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11659  GCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3929  GTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11714  GTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      3989  CTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11774  GTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      4049  CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11834  GTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      4109  GAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11894  GAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      4169  GATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      11954  GATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Qy      4229  GCGCGGAGCGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG
Db      12014  GCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG

```


QY	4289	GTGGTGAACCGGGTTGGCTCCCTGGTCTGGGGGAAAGCGCTCGAGGGAGGTGAGCCCAAT	4348
Db	12074	GTGGTTGGCGCGGGTTCCGGTCTGATGACCTCGGGAAAGCGCTCGAGGGAGGTGAGCCCAAT	12133
QY	4349	GGCGCGCGCGCACTCGCTCATCTTGGCCGTTGGTGGTGGCCGATCTGGTGAACAACCTTGTCCGG	4408
Db	12134	GGCGCGCGCGCGCTCGCTCATCTTGGGGTTGGTCTCCCGCGGGCGGGCTGGCGCGGGCGAG	12199
QY	4409	GCCGATGCGGAAGTTTGGTCATGGCGCCGGATCCGTTCCGGCGAGCGAGCCGTCTGGTGAAC	4468
Db	12194	GTCGAAGCGGAAGTTTGGAGGGGGCGGAGATCCGGGCGGGCAGAGGTGGCGCTGTCTGGTGAAC	12253
QY	4469	CACCGCCCGCGCGCTCGAAGGGGGGAGACCGGCTTGGTGGTGGGGAAGCTGAACAACCTCCGC	4528
Db	12254	GACGGCGCGCGCTCTGAAGGGGTGTTAGCGCGCTTGGTGGTGGCTGGAAGCTGAAGACTCCGCG	12313
QY	4529	GTCAACCGGATCCGCGCCACCGGACGTCACCCCGTCTGTCAGACCCAGGGCGTGGCGCGCGCTC	4588
Db	12314	GTCCCGGAGGGCTGCGCGGGCGGGCGCGCGCTGACACCGCGCAACCGAGGGCGTGGCGCGCTC	12373
QY	4589	GAAAGAAGACTTGAACCTGGTGGTGGGCGGGCGAGTCTTCGCGAGCGCCTTCAAGGTCTGG	4648
Db	12374	GAAATGACAGCGCGCAGCGCGCTGTCTGTCTGGCGAAGCTTCCGAGCTGGTCCGCGCGGAGGG	12433
QY	4649	TCCGCGCCCAACAGTGTACCGCCGACAGATCGCGCGCGGTCTGGCGGGGTGAACAAGCCGGCGAC	4708
Db	12434	GCGGCGCCCAAGAGGTGAGACCGCGCAAGAGAGCGCGGAGGTGGGTGTGACCCCGCGCGCGAC	12493
QY	4709	GTGGTCCGGGTGACAGCAAGACCGGTCGCGGGGTGACGTCGCGAGAACCGGGTGTAGTCC	4768
Db	12494	CTGGTCCGGGTGACAGGTGGCCGGTGTCTGGGTGATGTCCGCCAAGACCGGGGTGAAGCC	12555
QY	4769	GAGCGACCTCGCGCGGTGGCGCGGTGGCGCGGAAAGTTCATTGAACGACATGATCACTTCACC	4828
Db	12554	GATCAAGCGAGTGGTGGGGGGGTGGCGGAAAGTCAATGACGGCATGATCACTTCGCG	12613
QY	4829	GGTACGTCACCGGCGCGCGCAACCAATTCAGAGGAGCGGTGGCGTTGACGGTGGCGAAT	4888
Db	12614	GGTAGAGCGCGCGCGCGGTGGCGCGAGAGACTGAAGCCCGGCGGTGGCGTGTGACGTGGCGAC	12673
QY	4889	GCACTGCGGTACCCCGCAACAGTGGGCGAGACCGCGGGCTCGAATCTCCGCAACAGGGGGCCG	4948
Db	12674	GGCATGCGGAACCCCGCGAGCCCGGAGCGCGCTCTTGAACTCGCGGACAGCGGGGCG	12733
QY	4949	GCCGTTGGTGAAGCCAGTGTGTGTTCAGAGGCGCACTCGAGCGGGCGCAAGAACCGCTGGCGG	5008
Db	12734	GCGGTTGGAACAGCCACTGGCGTGTGAGAGGGCCCGGTGAGGCGCTCTGTACAGCCTTGGCGGG	12793
QY	5009	GTCCGCCATGTGTCGGCGCGCGCCACAGTGCAGAGGGGTGACAGAACGCTTCGAGACCGCGGAA	5068
Db	12794	GTCGATCGGGTGGTGGCGCGCCCAACAGAGACCGGCTGGTCAAAAGCGCGGGCGCGCGAA	12853
QY	5069	GATCGCAGATCGGTCGGTACGGCTTACAG	5099
Db	12854	GAATGCAAGGTGGATGAAGCGCTTTCACG	12884
RESULT 7			
US-10-398-605-3			
; Sequence 3, Application US/10398605			
; Publication No. US20040161839A1			
; GENERAL INFORMATION:			
; APPLICANT: Liu, H.			
; APPLICANT: Sherman, D.H.			
; APPLICANT: Zhao, L.			
; APPLICANT: Regents of the University of Minnesota			
; TITLE OF INVENTION: Method to alter sugar moieties			
; FILE REFERENCE: 600,475US1			
; CURRENT APPLICATION NUMBER: US/10/398,605			
; CURRENT FILING DATE: 2003-04-04			
; PRIOR APPLICATION NUMBER: PCT/US01/31255			
; PRIOR FILING DATE: 2001-10-05			
; PRIOR APPLICATION NUMBER: US 60/238,185			

[illegible]

Best Local Similarity 63.2%; Pred. No. 1.5e-268;
Matches 1979; Conservative 0; Mismatches 1138; Indels 15; Gaps 4;

QY 2395 GCGGATCAGGCGCCGATGCTGATCCGACCGGTGCGGACGTCCTCCGATTCGAGTGTGAG 2454
DB 55120 GCGAGGAAACCCCGTGTCTCCCGACCATCGGCGATCGTCTTCATCTGTTCAG 59061
QY 2455 CCGGCGGTGACCGATCTGTGAAACCGTGAAGAAATCTGTCCTCCGCTGCGGATC 2514
DB 59060 CCGCGGTGACCGCTGTGTGAAATCTGTCGAGAAATATTCGTGCGCAGGCTTCGAG 59001
QY 2515 GATGCTGCGGCGCGATGACGAAATCTGTGACGACCGAGTGCAGGCTTCGCTCGGAGT 2574
DB 59000 CACCGATCCCGCTGTGTGACGAACTGTCCACCACTGCGGCCAACGTGCGCGGAGA 58941
QY 2575 GACCGGCGCGGACGATGAGGCGGTGCGCCCGGACGCGCGGAGAAACCGGCTTCGAGT 2634
DB 58940 TACCGGTGATCGATGAGGCGGTGCGACCGGCGACCGGAAACCGCCCTTCCTCATG 58881
QY 2635 CAGGTAGACGTCTCCGACGATGATCTGACCGGACCTGCGGATGAGGCGGCGG 2694
DB 58880 GAGGTAGACATCACCGAGAAATCTCACCTGACCGACCTGCGGCGTGGGCGGACGAGG 58821
QY 2695 CATGCTTTCGCGCGGATCCGCAACAGCTGAGGCTGCGACCCCGACGAGGCTTCGAA 2754
DB 58820 CATGCTCTCGGCGCGGATACGACGAGCTCCAGTCTCCGCTCATGAGGCTGTGAG 58761
QY 2755 CCGGTAAACCGAGTCCGTCTGCAATGCGCGGCGTCCGTGCGCGCGGATGTCAGAACG 2814
DB 58760 GGGGTAGCCGTATGCAATGCAAGGCGGCGGTGCGGTGCAATGCTCTCCGGAACCG 58701
QY 2815 GCGGAGGACCTCTCCGCACTCGGTGCGCTGCGCGGCAACCTGCGGTGCGGACG 2874
DB 58700 GTGCACTCGGCGCTGAGAGCTCGGCGGCTGTCTCAGGAGAGCTTCCGTCGCGGCGCC 58641
QY 2875 GCTGATGCTCTGCGGACGCTGCAAGTGCAGCGCGCGGTGCGGCTGTGAG 2934
DB 58640 GCTGATGCTCTCGGAGGTTGATGATGATGATGATGATGATGATGATGATGATG 58581
QY 2935 TTGCGGATGAAATTCACACGAGTGCATGAGGCGGTTGCGCGCGCGGATGATG 2994
DB 58580 CTCGGGAGCAAAATTCATGACGCGCTGACGCGCGCGCGCGCGCGCGCGCGAGT 58521
QY 2995 GTTGAAGCGGACCGGACCGGTGCGCGGCTGCGCGCGGATCGGAGAGTCCGAG 3054
DB 58520 GCTGAAGCGGACCGGACCGGCTGCGCGCGGCTGCGGACCGAGTGTGAGAACCGG 58461
QY 3055 GTTCTTCTGACCGGTTGAGAGGCGGACGCTTGCAGGTGCGGCTGTGATCTGCTT 3114
DB 58460 GTTCCGCGCGGATCCCGGATGAGGCGCGGCTTGCAGGTGCGGCTGTGATCTGCTT 58401
QY 3115 GTTCAAGCGGATGAGAGGATGACGCGCGGAGGTCCACAGGCGGCGGTGCGGCG 3174
DB 58400 GTTCAACCGGATGAGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 58341
QY 3175 CAGCGTGTGTTGAGGAGGCGGAGGCTGTGATGAGGATGAGGAGGCGGAGGCGG 3234
DB 58340 GAGCTTGTGCTCTGTAAGGCGGAGGATGATGATGATGATGATGATGATGATGATG 58281
QY 3235 GCGGCGGTGACCAACGCGCGGATGACCGGATGATGATGATGATGATGATGATGATG 3294
DB 58280 GCGGCGCGGATGACCAAGTGCACCGGCGGATGATGATGATGATGATGATGATGATG 58221
QY 3295 GAGGTATCATGCGCGGTGCTTCGAGCGGACCTGCTGATGATGATGATGATGATGATG 3354
DB 58220 GAGGTATCATGCGCGGTGCTTCGAGCGGACCTGCTGATGATGATGATGATGATGATG 58161
QY 3355 GTTGAAGGAGGACCGGATGCGGCGGTGATGATGATGATGATGATGATGATGATGATG 3414
DB 58160 GTTGAAGGAGGACCGGATGCGGCGGTGATGATGATGATGATGATGATGATGATGATG 58101
QY 3415 ACAAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3474

DB 58100 ACAAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58041
QY 3475 CCGCGGATCCGACCGCTTCACACCGCGCGGCGGCTTCAGGAGGCTGATGATGATGATGATG 3534
DB 58040 CCGGTGAGGACCGGCTGCAAGAGCGCGGCTTTCGAGGCGGAGGACCGGTGATGATGATG 57981
QY 3535 GTACTTGCAGCGGCGGCGGCTTCGACCGCGGCGGCGGATGATGATGATGATGATGATGATG 3594
DB 57980 GTAGGACCGGAGCGGCTGCTTCACCTGCGGTGCGGACCGGAGATCCGCGGAGGAG 57921
QY 3595 GTGAGCAGCTGCGGACCGGCGGCGGCTTCAGGATGAAACCGGCGGATGATGATGATGATG 3654
DB 57920 GTGAGCAGGCGGCGGACCGGCGGCGGCTTCAGGATGAAACCGGCGGATGATGATGATG 57861
QY 3655 CCGGAGGAAACCGGAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATG 3714
DB 57860 CCGGAGGAAACCGGAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATG 57801
QY 3715 TTGCGGCGGTGCGGATGCGGACCGGATGAGGCGGATGATGATGATGATGATGATGATG 3774
DB 57800 CCGGATCTGCGCGGCGGACGATTCGATCCGCGGACCTGCTGATGACGACGCGGCG 57741
QY 3775 CACGCGCGGCGGATGCGGCTGCGGTGCGGCGGCGGATGATGATGATGATGATGATGATG 3834
DB 57740 CCGCGGCGGTGCGGCTGCGGCGGCGGACGAGGCGGCGGATGATGATGATGATGATGATG 57681
QY 3835 CCGCGGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3893
DB 57680 CTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57621
QY 3894 CACGAAACCGGTGCGT-----TTCCGCTGTCGCGACCTCGGCTGATGATGATGATGATG 3945
DB 57620 CACGCTGTGCTGCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57561
QY 3946 TGGTGTGCGGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4005
DB 57560 GGGAGACCGGCAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57501
QY 4006 GCGCGGTGCGGAGGCGGAGGACCGGTTGCGGCGGACCTGCGGTGATGATGATGATGATG 4065
DB 57500 AGCGGCTGCGGAGGCGGAGGACCGGCGGCGGAGGCGGCTGCGGTGATGATGATGATG 57441
QY 4066 GTGCGCTCGGTGCGGATGAGGCGGATGATGATGATGATGATGATGATGATGATGATG 4125
DB 57440 GCTGCGGCGGACCGGATGAGGCTTCATGATGATGATGATGATGATGATGATGATGATG 57381
QY 4126 CGACCTTCTCCGCGGACGATGCGCTGAGCGGCTGACGATGATGATGATGATGATGATG 4185
DB 57380 CGACCTTCTCCGCGGACGATGCGCTGAGCGGCTGACGATGATGATGATGATGATGATG 57321
QY 4186 CGTCAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4245
DB 57320 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57261
QY 4246 CGGTGACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4305
DB 57260 CGGTGACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57201
QY 4306 GCTCCCTGATCTCGGCGGACGATGCGGATGAGGAGGATGAGGAGGCGGCGGACCTGCG 4365
DB 57200 GCGCGGCGGACCTCGGCGGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 57141
QY 4366 TCACTTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4425
DB 57140 TCACTTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 57081
QY 4426 GATGAGCGGAGTCCGCTTCCGACGAGGCGGATGATGATGATGATGATGATGATGATGATG 4485
DB 57080 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 57021
QY 4486 AGGCGGATGACCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4545
DB 57020 AGGCGGATGACCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 56961

QY 4546 CCGAGCTCCACCCGCTGTCAGACCCAGAGGCGTGGCGGCTGGAAGAAAGCTTACCT 4605
DB 55960 CCGAGGCGCCCTGTGAGGTGACCCAGTGGCGGTGGGAGGTGAAGAAAGAGCGCGAGC 56901
QY 4606 GGTGTGCGGCGGAGTCTTCCGAGGCGCTCCAGAGTGTGGTCCGCCCCAGAGGTGA 4665
DB 56900 CCGGCTGCGGCTCAGCTTCTCCAGCGCGTTGAGAGTCCAGGCGGCGGCCCCAGAGGTGA 56841
QY 4666 CCGCGAGTCCGCGCGCTCTGCGGGGTGACGAGCGCGCGAGCGTGGTCCGCGTCCAGCA 4725
DB 56840 CCGCGAAGACCGCGAGGTCGCGCGCGCTGATGGCGCGCGACCGCGACGCTGTGCGAGCG 56781
QY 4726 GACCGGTGCGCGGCTCCAGCTCCGAGAACACCGGTGTGATCCGAGGCACTGCGCGCGT 4785
DB 56780 ATCCGCTGCGCGGCTGATGTCCGAGAAAGACCGGCTCCAGCGCCGAGCGACGCGCGT 56721
QY 4786 GCGCGGTGCGCGGAGGTGATGACGCGCATGATCATCTTCCAGCGTGAAGTCAACCGCGC 4845
DB 56720 GTGCGGCTGCGGCGAAGGTGAGCGATGATCATCTTCCAGCGTCAAGCTCGGTGGCGT 56661
QY 4846 GCAGCACCACTTCCAGGCGAGCGGTGCGGTGCGAGTGGCGAGTCCGCTGACCCGCA 4905
DB 56660 GCGCGAGACGAGTGAAGGCGCACCGTGGCGTTCAGTCCGACGCGAGTGGCGACCCGCA 56601
QY 4906 CCAAGTGGCGGACCCCGGCGCTCCGAACTCCGCGCACAGGCGCGCGCGGTTGGAGCGAGT 4965
DB 56600 CCAATGTGGCGGACCCGCTCCGTAAGTCCGTCGACAGCGCGCGCTGTGTGTCAACACT 56541
QY 4966 TGTGTTCAGGCGCCCACTCCAGGCGGCGAGGAAACCGCTGCGCGGTGCGCGATCTCGGCG 5025
DB 56540 CGTGTGTGAGGCGCCAGAGACGCGCGTGAACAGACTGCGCGCATTCAGTGGGCGC 56481
QY 5026 GCGCCAGGTGAGGCGGCGTGAAGAAACGCTCGGAGCGCGCGAGATGCGCATCTCGGTCG 5085
DB 56480 GCGCGAGTGAATCTGCTGAGGAGAAAGTGGGTCTCCGAGAGTGCAGATCCCGCA 56421
QY 5086 GTAGCGGCTTCA-CGCGCTCGCGCGGTGTGAGACGCGCGAGCGCGCGAGCGAGGCTGC 5144
DB 56420 GATCTCGCTTCAATGCGCTGCTCTCCCTTACGCGCGCACAGCTTCCGCGAGTGGCG 56361
QY 5145 GCAGCTGATGTGTAAGTGAAGTGGCTGTG-----CGCGAGCATTTCCGATCTGCGCGAA 5199
DB 56360 CCGCGGCGGCGGCGAGCGCTCAGGTCTGCGGTACGCGCTGAGCTCGGAGCGAGTGGCGG 56301
QY 5200 GGTCAATCCACCGGTGTGCGGCGGACCGCGCTGTGAAAGTCCGCGGCGACCTGACGAG 5259
DB 56300 GCGCACCGCGCGCTGCAAGGAGTGCCTGCGCGCGTGTGCTCTGCGCGCACACCTCGCG 56241
QY 5260 CATGTACCGGTCTCTGTGCGGTGAGAACCGCGCACCTCTCTCGAGTGCACAGGCGTGTGA 5319
DB 56240 GTGCGCGTGCACCGCGGCGCACAGGAGTGCCTCGCGAGTCAAGAGTCAAGACCTCG 56181
QY 5320 GCGGATCTGGGTACCGGAGACGTCAGACGATGTCGAGTGTGTCGCGGCTGTCTCTC 5379
DB 56180 GACCTCCCGGTGCGGAGCGCGGCTTCCAGGAACTGCCAAGCGTGTGATGTCCTGACC 56121
QY 5380 GCGGACGTGCGGTGATGTGTGCGGTGACAGTGCACCGTCCGCGCAACTGCGCGAGTGC 5439
DB 56120 GGTGCGGCGCGAGAGAGAGTGTGCGGCTCGGCTCCGCGCAAGCGCGCGAGCGGCG 56061
QY 5440 GTGCGCACCGCGCTCGGTGCGAGGTGACAGGCGCGGTGAGAGGCGCGTGCATCTCTT 5499
DB 56060 CCGGCGTGGCGGCGGTGTGTGATCAAGAGTTCGCGCACTCCGCGCGGTCCAC 56001
QY 5500 GACGAGAGGCGC 5511
DB 56000 GTCCGAGTAC 55989

RESULT 9
US-10-229-148B-1
; Sequence 1, Application US/10229148B

Publication No. US20040091975A1
GENERAL INFORMATION:
APPLICANT: Meiji Seika Kaisha, Ltd.
TITLE OF INVENTION: Midcamycin biosynthetic genes
FILE REFERENCE: 138451 US
CURRENT APPLICATION NUMBER: US/10/229,148B
CURRENT FILING DATE: 2002-08-28
PRIORITY FILING DATE: 2002-07-19
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 84428
TYPE: DNA
ORGANISM: Streptomyces mycarofaciens
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (1) .. (675)
OTHER INFORMATION: ORF42 (fragment)
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (1168) .. (2202)
OTHER INFORMATION: ORF41
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (2220) .. (3215)
OTHER INFORMATION: ORF40
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (3237) .. (4691)
OTHER INFORMATION: ORF39
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (4695) .. (5948)
OTHER INFORMATION: ORF38
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (6048) .. (6629)
OTHER INFORMATION: ORF37
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (6653) .. (7945)
OTHER INFORMATION: ORF36
FEATURE:
NAME/KEY: CDS
LOCATION: (8149) .. (9015)
OTHER INFORMATION: ORF35
FEATURE:
NAME/KEY: CDS
LOCATION: (9012) .. (9335)
OTHER INFORMATION: ORF34
FEATURE:
NAME/KEY: CDS
LOCATION: (9328) .. (10458)
OTHER INFORMATION: ORF33
FEATURE:
NAME/KEY: CDS
LOCATION: (10521) .. (11603)
OTHER INFORMATION: ORF32
FEATURE:
NAME/KEY: CDS
LOCATION: (11729) .. (12961)
OTHER INFORMATION: ORF31
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (13016) .. (14044)
OTHER INFORMATION: ORF30
FEATURE:
NAME/KEY: CDS
LOCATION: Complement (14074) .. (15096)
OTHER INFORMATION: ORF29
FEATURE:
NAME/KEY: CDS

LOCATION: (15643)..(17466)
 OTHER INFORMATION: ORF28
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (17522)..(18895)
 OTHER INFORMATION: ORF27
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (19063)..(20229)
 OTHER INFORMATION: ORF26
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (20307)..(21743)
 OTHER INFORMATION: ORF25
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (21733)..(22527)
 OTHER INFORMATION: ORF24
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (22534)..(23571)
 OTHER INFORMATION: ORF23
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (23555)..(24463)
 OTHER INFORMATION: ORF22
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (24460)..(25650)
 OTHER INFORMATION: ORF21
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (25647)..(26105)
 OTHER INFORMATION: ORF20
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (26180)..(27391)
 OTHER INFORMATION: ORF19
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (27937)..(28983)
 OTHER INFORMATION: ORF18
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (29244)..(42779)
 OTHER INFORMATION: ORF1
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (42823)..(48657)
 OTHER INFORMATION: ORF2
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (48712)..(59802)
 OTHER INFORMATION: ORF3
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (59850)..(64556)
 OTHER INFORMATION: ORF4
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (64687)..(70365)
 OTHER INFORMATION: ORF5
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (70365)..(71078)
 OTHER INFORMATION: ORF6
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (71113)..(72360)
 OTHER INFORMATION: ORF7
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (72400)..(73665)

OTHER INFORMATION: ORF8
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (73694)..(75043)
 OTHER INFORMATION: ORF9
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (75899)..(76570)
 OTHER INFORMATION: ORF10
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (76602)..(77765)
 OTHER INFORMATION: ORF11
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (78039)..(79313)
 OTHER INFORMATION: ORF12
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (79391)..(81052)
 OTHER INFORMATION: ORF13
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (81541)..(82356)
 OTHER INFORMATION: ORF14
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (82760)..(83362)
 OTHER INFORMATION: ORF15
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Complement (83495)..(84142)
 OTHER INFORMATION: ORF16
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (84329)..(84428)
 OTHER INFORMATION: ORF17 (fragment)
 US-10-229-148B-1

Query Match 13.5%; Score 1203.2; DB 7; Length 84428;
 Best Local Similarity 65.5%; Pred. No. 5e-260;
 Matches 1778; Conservative 0; Mismatches 933; Indels 5; Gaps 1;

QY	2387	CAGACGCGCGCGATCAGGCGCCCGATGCTGATCCACCCGTCGCGACGTCCTGGAAT	2446
DB	3232	CAGGCTTATCGAAGAAAGCGCGCTGCTCGCCCAACCCGCTATCGACTTGGAGC	3291
QY	2447	TGGTTGAGCGCGCGGTCACGACCTGTGGAACGCTGAGAAAGAACTGTCGCCGCG	2506
DB	3292	TGGTTCAGCGCGGTGACGACCTGTGGAACCGTCCATGAAAGTACTGTGCGCTCG	3351
QY	2507	TGCGATCGATGCTGCGCGCGCGACGACGTAAGTGTGACGACCGAGTGACGAGCTCG	2566
DB	3352	GCGGAGGAATCTGCCACCGTGGTCAAAAAGCGCTCAACACCTGGGTGAGCGTGT	3411
QY	2567	TCCGCGGTATACCCGCGCGCGACGTAAGCGGTGCGCCCGCGACCGCGGAAACCGG	2626
DB	3412	CCGCGGGAACCTTGCCTGCTGAGCGGTGCGCTCCCGGACGTCGCCGAAACCGG	3471
QY	2627	TCCCGGTACAGGTAGACGCTCCGACGAGTGTGATCTGACCGCGACCTGGGATGGG	2686
DB	3472	TGCGGTAGAGGTAGAGTGGCGCCAGCAATGACCTGACCTGCGACCTGGGATGCG	3531
QY	2687	GTCGCGCGCATGTTTCCGCGCGATCCGCAACAGCTGCGCTGACACCCCGACGACG	2746
DB	3532	GTCGCGCGCATGTTTCCGCGCGATCCGCAACAGCTGCGCTGACACCCCGACGACG	3591
QY	2747	CTCTCCACGCGTAAACCAAGTGTGCTGACGCGCGGATCCGCTCGCGCGCTAGT	2806
DB	3592	CTGTTCAGCGCGTACCGCTGAGTGAAGTGAAGGAGTGTCTTGGCGGACCTTCTCC	3651
QY	2807	ACGAACCGGCGAGGCGCTCCCGCAACTGCGTCCGCTGCGACCAACTGCGCTG	2866

Db 3652 CGAAGCCGGTCAGCGCTTCCTGAGAGTCGCGCGCTCGGCAACCGGACAGAGAGCCCTGCG 3711
QY 2867 TCCCGACCGCTGATAGCTCTCGCGGACCGGTACGAAGTCGAGCGGCGGTGGCGGCTGAC 2926
Db 3712 GCGCGGCGCGTGAAGCTCTCAAGGAGGTACGAAGTCAGTGGTCGGTCGCGCGGCTG 3771
QY 2927 TCGTTGAGTTCCGAGTGAAGTCGACCAAGTCAGTGAAGCGGTGAGCGGTCGCGCGGCGG 2986
Db 3772 GCGTTCACTCGCGGATGAAGTCGACCAAGTCAGTGAAGCGGTCGCGCGGCGGCGGCGG 3831
QY 2987 ATGATGAGTTGAGGCGGACCGGATCGCGTCCGCTTGGCGCGCATTCGACAGAG 3046
Db 3832 ACGATGATGTTCAAGCGCGAGTTTCAACCGTTTCGCGCGTGGGAAACCGACCGCTGAGAG 3891
QY 3047 TCCCGAGAGTTCTTCTGACCCCGTTCCAGAGCGGCAACCGTTCCGCGTGGTCCCTGGATC 3106
Db 3892 CGCTCCAGAGTTGGCTTTCACCCGTCGAAAGCGGCTTCTTCTGAGTGGCGCGAT 3951
QY 3107 TCGTCGTTGTTCAAGCCCGTACAGAGAGGTACGACCCGCGCGGAGTCCACAGGCGCGGCG 3166
Db 3952 TCCGCACTGTTCAACCCGTCAGAGAGGTCCGCAACCGGTGCAAGCGCGCAACAGGCGCGCC 4011
QY 3167 TGGCGGCGGACGCTCTGTTCCGTGAGGCGGAGGCGGTTGTTGTAAGTCGTCAAGGCGGAG 3226
Db 4012 TGCCTTTCAGCGCTCCGGTCGGTAAAGCGGAGCGGTTGGTGTACAAAGGTGAGCGGAGAT 4071
QY 3227 CCGCGCGGCGGCGGTGCGACCAACGCGCCCGATACCCCGGTTGTTCAAGGTTCCAG 3286
Db 4072 CCGCGGCGGCGGCGCTTCGAAACGAGGAGCCGAGCCGGAATGGTAGCGGCTCCAGA 4131
QY 3287 CCGCGGAGAGGTACATCGCGCTCGGCGGTTCTCCAGCGGACCTGTCAGTACCGAGTG 3346
Db 4132 CCGCGCGGACAGTACATGCGCTCGGCGGTTGTCGTTGCGGACATGTCGTGATAGAGCGCGCG 4191
QY 3347 AGGATCTGTTGCGGCGGACCAACGCACTCCGCGTGTACTTGGCACTGTCAACCGGTACG 3406
Db 4192 AACATCTTGTGCGCTCGTCCAGGCGGAGTGTATGTAGCGGCGCGCGTACCCGTTACG 4251
QY 3407 CAGAAATGACAGCGGAAATGACACTCGGCGCGGTTAGAGCCGCTGATAGCGGAG 3466
Db 4252 CAGAAATGACAGCGGAAATGACACTCGGCGCGGTTAGAGCCGCTGATAGCGGAGAC 4311
QY 3467 GCGGCGCTCCGCGGTACCGGACGCTCCAAACGCGCGGCGGTTGAGCGGCGCTGATCGTG 3526
Db 4312 ACCGCTTGTGTTGTAAGCGGCGCTGAAAAGCGCTTGGCTTCCAGCGGAGCAAGGTTG 4371
QY 3527 TTGCTCAGTACTTGGCGCGGCGCTGCTTCACCGCTGTCAGCTCGCGGATGCGCA 3586
Db 4372 TTGCTCAGTACTTGGCGCGGCGCTGCTTCACCGCTGTCAGCTCGCGGATGCGCA 4431
QY 3587 TCGAACAGTGAAGCTCGCGGACCCGCGCGGTTCAAGTTCGAACCGAGCGGATC 3646
Db 4432 CCGAAGAGTCAGAGAGCGGAGGAGCAATCGGTGACTCTTGAATTGAGAGGCGC 4491
QY 3647 TGTCTCAACGAGGTGAAGCGGACGAGCGGTGATGAGGCGGTTGAGTTCGCGGAGCGGAC 3706
Db 4492 TTTTTCAGGAGTGAAGGAGCGGCGCGCTGATGATCCGAGCGAGCGCTACCAATGCGCG 4551
QY 3707 GCGTGTGTTCCGCGCGGTGTCGATCGGACCGGATGAGGCGGTTGACCTTCGCGCTGAGC 3766
Db 4552 GCGGCGTCCGCGCGCATTCGCGCTCCGTCATATGCGGAGGTGACCAAGTTCCGCGGAGA 4611
QY 3767 GCGCGACACAGCGCGGCGGTCGCGGTGCGGTGCGCGCTCGCGCATCTCCGTCCGT 3826
Db 4612 GCGTCGAGAGCGGCGCGGATCACTCGCGGAGAGGTGACAGCGCTTACGAGTGTTCGC 4671
QY 3827 ACAGCGGTGCGGTGAGAGTGTCTTCAATCGAGTGCAGAACCTTCTGGCGCTGAGTG 3886
Db 4672 ACCGCTTCTTGGCTCAACATGAGTATCCACCTCTTCCATTGCGAAATGCTCTTGATG 4731
QY 3887 CGAGATCA-----CGAACGTTGCGTTTCGCTTGTCCCACTCGCGCTGATCAACGA 3941
Db 4732 AGAAGGCGGTGCGGTGATCACTGAGAGTCCGCTGACGCGGCGGTCAGCTCGGCG 4791

QY 3942 CCGCTGAGTGGCGAGTCGATGATGATGACACACCCGCGGATGTTCTTCACTGAGACCC 4001
Db 4792 CCGCGGTTGGCGGCGGACCGGACGATGTCATACCCGCGGATATCTCATGGAGAGCC 4851
QY 4002 GAGGCGCGGTCCGAGAGGCGGAGACCCGTTCCGCGGACCACTTCGCTCTTGTTCAGCGCG 4061
Db 4852 GTGAGACCGGTCCGAGAGGCGGATGACCCGCGCGGAGAGGCGCTCGGTGCGGAGGTGC 4911
QY 4062 AGCGGTGCTCCGTCGAGTGGGCGGTCACTGATGAGCGGAGGAGAGTGAAGGCTGT 4121
Db 4912 GCGTTCCGACGGAACCGGTACGCTCCAGTTGGTGGAGGCGCGGAGAGTGAAGGCTGG 4971
QY 4122 GCGACGACCTTCTCCGCGCGAGAGTCCCTGACCGGTCAGCGTCAATGCCGTGGCG 4181
Db 4972 GCGACGACGTTTTCGCGGAGAGGCGGAGGAGGAGGAGTCCGATGAGACCCGCTGAC 5031
QY 4182 GTGCCGTCACAGAGTATCAAGTACTGATGTTGCTCTCTCTGTCGCGCGGAGAGAG 4241
Db 5032 TCCCTGTCATGTCAGACAGATACGATGATGTTGCTCGTTCTCATGCAATGCG 5091
QY 4242 TGCACGAGTGAAGCGCGGTAGTGGCGGAGCTCGCTGATGAGCGGCTGATGAACCGG 4301
Db 5092 AAGACGCGCACACCGGTACTCCGAGAGTTCCGTGCGGTAGTCTATAGTTGCTTGG 5151
QY 4302 TTGTCCTCCCTGCTCCGCGGAAACGCTCGAGGAGGTGAGCCCATGAGCGCGCGCAC 4361
Db 5152 TTGTGCGGTACGACTTCTTCAACAGCTCGAGGAGCGTCAAGCCCATGCGCGCGAGGCT 5211
QY 4362 TCGCTCATCTTGGCGTTGTTGTCGATTCGATGACCACTTGTCCGCGCGGATGCGAG 4421
Db 5212 TCGCTCATCTTGGCGATGTTGCTCCCGCGGAACTGACTTCTTCAGGCGGAGTCCAG 5271
QY 4422 TTGTGATGAGCGCGGATCCGTTGCGGAGGAGCGGCGGTGTCGAGTCAACCGCGCGCC 4481
Db 5272 TTGTGAGAGAGCGGACACATGAGCCAGCTCGTGTGATGAGTGAACCGCACCGCC 5331
QY 4482 TCGAAGCGGTGACCGGCTTGTGTCGAGTGAAGTGAACACTTGAGGCTCACCGGATCCG 4541
Db 5332 TCGAAGGAAATTGACGACCTTGTGCGGTGAAGTGAATACCTCGGCGTCCGAAACGG 5391
QY 4542 CCGACCGGAGCTCAACCGGTGTCAGCCGAGGCGGTGAGGCGGTGAGAGAAAGCTTG 4601
Db 5392 CCGACAGGCTTCTCGTCGAGTGTCTCGATCCGATGCGGTGCGGAGAAACAGCGCG 5451
QY 4602 ACCGTGATGTCGAGGCGGATCTTGCAGAGCGCTCAACAGGTGCTGATCGGCGCCACAG 4661
Db 5452 ATGCCGCTTCGCGGCGGACGCTCTTCAAGCGCTGACATCAACGCGCTTACCCACAGA 5511
QY 4662 TGTACGCGAGATCGCGCGGTCTGCGGAGTGAAGAGCGGCGGACGATGTCGCGGTG 4721
Db 5512 TGCACCCGAAATTCGCGAGGTGCGGCGGTGATGAGGCGCGCACCGCTCGGAGATCC 5571
QY 4722 ACCAGACCGGTGCGCGGTGCACTTCGAAACACCGGTGATGATCCGAGCACTCGCC 4781
Db 5572 ACAATCCGATGATGATCGATCGCGAAGACCGCTCCAGTCCGAGCGACCGCACT 5631
QY 4782 GCGTGGCGGTGCGCGGAAAGTCAATTGAAGGCAATGATCACTTCAACCGGTGACGTC 4841
Db 5632 GCGTGGCGGTGCGCGGAAAGTCAAGCGCGGCAATGATCACTTCAACCGGTGACGTC 5691
QY 4842 GCGCGAGACCAAGTTCCAGGAGGACAGTGGCGTTGCAAGTGGAGATGCAAGTGCCTTAC 4901
Db 5692 GCGTGTACCAATGATGATGAGGCGGACGATGCGATTCGAGAGTGCAGAGGATGAGACC 5751
QY 4902 CCGAACAGTTCGCGGACCCGCGGCTTCAATTCGCGCACAGGAGCGCGCTGTTGATGAC 4961
Db 5752 CCGGCGAATTCGCGGACCCGCTTCTCGAATCTCGAGGTCAAGGAGTCCCGCTTGTGAGG 5811
QY 4962 CAGTTGTTGTTCAAGGAGCCCATTCAGAGGCGGCGAGAAACCGCTGCGGTGCGCATGCTC 5021
Db 5812 CACTGTGTGTCAGGCGCCAGGTGACGCGGTGCAACACCGGAGAACGATGATGAGATTTC 5871

QY 5022 GCGCGCCGACGTCAGAGGGGTGAGAAACGCTCGGAGCCGCCGAAGATCGCCAGATCG 5081
 DB 5872 GGGGGGCGCCACAGAAAGCGCTGAAGAAAGTTGGCGGCGCCCGGAAACGGCCAGATCG 5931
 QY 5082 GTCGGTACGCGCTTCA 5097
 DB 5932 CCGAGTTCGCGCTTCA 5947

RESULT 10

US-09-793-708-24/c
 ; Sequence 24, Application US/09793708
 ; Publication No. US20030104597A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ASHLEY, Gary
 ; APPLICANT: BETLACH, Melanie C.
 ; APPLICANT: BETLACH, Mary C.
 ; APPLICANT: MCDANIEL, Robert
 ; APPLICANT: TANG, Li
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 30622002121
 ; CURRENT APPLICATION NUMBER: US/09/793,708
 ; CURRENT FILING DATE: 2001-02-22
 ; PRIOR APPLICATION NUMBER: US 09/657,440
 ; PRIOR FILING DATE: 2000-09-07
 ; PRIOR APPLICATION NUMBER: US 09/320,878
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: US 09/141,908
 ; PRIOR FILING DATE: 1998-08-28
 ; PRIOR APPLICATION NUMBER: US 09/073,538
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: US 08/846,247
 ; PRIOR FILING DATE: 1997-04-30
 ; PRIOR APPLICATION NUMBER: US 60/134,990
 ; PRIOR FILING DATE: 1999-05-20
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 24
 ; LENGTH: 1565
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-793-708-24

Query Match 8.2%; Score 729.4; DB 3; Length 1565;
 Best Local Similarity 68.1%; Pred. No. 1e-153; Indels 0; Gaps 0;
 Matches 1015; Conservative 0; Mismatches 476;

QY 2369 GGGCGGGTCGACGCGCTTCAAGACGGCGCGATCAGGCGCCGATGCTGATCCCAACCGGTG 2428
 DB 1530 GGGGGGATGGGGCGGGGTGGGGTCAAGCGAGAGAACCGGGGCTCTCCAGCGGTCC 1471
 QY 2429 GCGACGTCCTCGTTGAGTTGTTGAGCGGGCGGTCAACCGATGTTGAAACCGTTCAGG 2488
 DB 1470 GCGCGGTGCGGCTTCAGCTGTTTCAAGCGGGCGGTGACGACTGATGAAAGCGTTCATG 1411
 QY 2489 AAGACCTCGTCCCCCGGCTCGCGATGATGCTGCGGCGCGACGCTGACGAAAGTGTGAGCG 2548
 DB 1410 AAGTACTCTGCGCGTTCAGCGCGCGACCTGCGCGCGCTTCGACGAAAGTCTTCGACG 1351
 QY 2549 ACCGAGTGCAGCTCGCTCGTCCGGGTGACCCGGCGCGCGAGCTAGCGGGTGGCGCCGCC 2608
 DB 1350 ACTCGGTGAGGAGGTGTGGGGGTCAACCGGCGCGCGAGTGAAGCGGGTGGCGGTCC 1291
 QY 2609 AGCCCGGGGAAACCGGCTCTCCGGTACAGGTACAGTCTTCGAGAGAGTGTGATCTGACCC 2668
 DB 1290 AGGTTCGGGAAAGCGGCTCTCGCGTACAGGTACAGTCTTCGAGAGATGACCTGACCC 1231
 QY 2669 GCCACCTGGGAGTGGCGGGTGGCGCGCATGCTTTCCGGCGCGGATTCGCGAACACTGGGG 2728
 DB 1230 GCGACCTGGGAGTGGCGGGTGGCGCGCATGCTTTGATTCGCGAACACTGGGG 1171
 QY 2729 TCGACACCCCGACGAGGCTCTTCAACGCGGTAAACCGAGGTGGTCTGCAATGCGCGGGTTC 2788

DB 1170 TCGGCCCCGGTGGCGAGGCTGTTCAAGGGCGTAAAGCCGTAAGTGTGAATCGCGGGGTG 1111
 QY 2789 CGCTCGGCGGTGATGACGAAACGGGCGAGGCTCTCCGCAATCGGTCTGTTGGCC 2848
 DB 1110 CGCTCGGCGAACCGGCTCTCGAAAGCGTTGAGGGGCTCTGAGACTCGGCGGCTCTCC 1051
 QY 2849 TCGGACAAACGTCGCGGTGTCCCAACCGCTGTAGTCTCCGCGAGCGGTGACGAACTTCAGC 2908
 DB 1050 TCGGACAGCTTGGCGGTGTCAACCGCGCTGTAGTCTCGGAAATGTGAACAAATTCATTC 991
 QY 2909 GGCCTGTGCGGCTGAGCTGTTGAGTTCCGCGATGAAAGTGCACAGGTTCGATAGGCGG 2968
 DB 990 GTCTGCGCTTGGCGGCGGTGTGAGGTCCGCAATGAAGTGCACAGGTTCGAGAGCGG 931
 QY 2969 TTGGCGCGCGCGCGAGATGATGTGGTTGAGCGCGAGCGGATCGGTCGCGCGTTCG 3028
 DB 930 GAGGACGCGCGCGGAGACGATGTAGGCGAAGCGGATGTGATCGCGACTCGCGCTCG 871
 QY 3029 GCGCGCATCGCGAGAAAGTCCCGAGGTTTCTCTGACCCGTTGAAAGCGCGCACGCTTG 3088
 DB 870 GCGCGAGCTGTGAAAGCGGCGAGGTTCTCGCGACGCGCGAAAGCGGCGCTTCTTG 811
 QY 3089 CCGGTGTGCGCTGTGATCTGCTGTTGTTCAAGCCCGTACAGGAGGTACGAGACGCGCG 3148
 DB 810 CCGGTGTGCTGTGATCTGCTGTTGTTGAGGCGGTGAGCGAGGTGCGATGCGTTC 751
 QY 3149 AGGTCCCAAGGCGCGGCTGCGGCGCGAGGCTGTTTGGTGTGAGGCGAAAGCGCTTGTG 3208
 DB 750 AGGCGCCAGAGGCGGCGGCTGCGCTCAAGGCTGCGCTCGGTGAGCGGAAAGGATGCTG 691
 QY 3209 TAGACGCTGACGAGGCGAAAGCGCGCGCGCGCGCTGCGTGTGAGGCGCGGCG 3268
 DB 690 TAGACGCTGAGCGCGAGCGCGGTGTGAGGCGCGCGCGCGCGCGCGCGCGCG 631
 QY 3269 TTGATCAACGATTCAGAGCGCGCGAGAGGTACATGCGCGGTTCGCGACGCGAC 3328
 DB 630 TTGATGAGCGGCTGACAGCGCGCGCGAGAGGTACATGCGCGAGGATGCGCGGATTC 571
 QY 3329 TCGTGTGATGACCGAGGTGAGGATCTGTTGCGCGGACCAACGACTCGCGTGTACTTG 3388
 DB 570 TCGTGTGATGACCGAGCGGAACATGCGGCTGCGCGTGAAGCGCGAGCGGTGTGCGG 511
 QY 3389 GCACTGTGACCCGCTACGAGAAAGTACAGCGAGCGGAACATGACGCTCGGCGGTGAGC 3448
 DB 510 GCGCGGTGACAGAGCGAGAAAGTGTGAGCGAGCGGAACATGAGGTGCGGCGGTGAGG 451
 QY 3449 CCGAGCTGTGACGAGAAAGCGGCGCTCGCGCGGTACCGGAGCTTCGCAACGCGCGCGCG 3508
 DB 450 CCGAGCTGTGACGAGAAAGCGGCGCTCTGTGCGAGCGCGCGCTGTGAAAGCGCGCGTGT 391
 QY 3509 TTGAGCGGCTGTGATGCTGTTGCTCAAGTCTTGGCGCGCGCGCGCTTCGACCGCGTGT 3568
 DB 390 TCGAGCGGAGAGCGGATGTTCTTCAAGTACGCGCGCGCGCGCGCGGTGTGACCGGCGT 331
 QY 3569 CCGAGCTGTGAGGTGAGTGTGACAGAGTGTGAGAGCTTCGCGCAACCGCGCGCGGTTCAGG 3628
 DB 330 CCGAGCTGTGAGGACTGTGCGCGGAACAGGAGGAGCGCGCGGAAGCGCTCTCCGCTGACG 271
 QY 3629 TCGAACCGAGCGGCGACTGTGCTCAACGCGGAGTGAAGCGGAGCGCGCGTGTGAGGCGGTG 3688
 DB 270 CCGAGGTGTGCGGCGGCTCTCTCAAGCGGAGTGAAGCGGAGCTGTGCGTGAAGCGAGCG 211
 QY 3689 AGTTTCGAGAGCGCGAGCGGCTGTGCTTGTGCGCGGTGTGAGATGCGGACCGGTGAGCGG 3748
 DB 210 AGCGGAGAGAGGTGCGGCGGCTGTGCTGCGGCTGTGCGGCGGAGCGAGGCGCGCGGCG 151
 QY 3749 GTGACCTGCGCGCTGAGCGCGCGCGCGACAGAGCGCGCGGTGTGCGGCTGTGCGCGG 3808
 DB 150 GCGAGGCTGTGAGCGAGCGGCTGTGAGCGCGCGCGCGCGAGATGCGGCTGTGCGGCGGAG 91
 QY 3809 TCGGAGATCTCGGTGCGGTACGCGGCGCGGTGAGAGTGTGCTTCATGAC 3859
 DB 90 CGTTGCGCGGCGGCGGTGCGAGAGGCGGCGGCGGTGTATCGGAGCGCTC 40

RESULT 11
US-10-201-365-22/c
; Sequence 22, Application US/10201365
; Publication No. US20030148469A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: COMBINATORIAL POLYPEPTIDE LIBRARIES PRODUCED USING A MODULAR
; FILE REFERENCE: 30062202103
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1565
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-201-365-22

Query Match 8.2%; Score 729.4; DB 6; Length 1565;
Best Local Similarity 68.1%; Pred. No. 1e-153;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY	2369	GGGCGGCTGACGCGCTTCAGACGCGCGGATCAAGCGCCCGATGCTATCCACCCCTCG	2428
DB	1330	GGGGGAGATCGGGCGGGGTGCGGGTCAAGCGAGAAAGCGCGGGCTCTCCACCGCTCC	1471
QY	2429	GCGAGCTCCCGTTCAGATGTTGATGACCGCGCGGTGACCGACCTGCTGTAACCGTCAAG	2488
DB	1470	GCGGCGTTCGCTCCAGCTGCTTCAAGCGGGCGGTGAACCTTGAATCCAGCCCTGCTCAG	1411
QY	2489	AAGAACTGCTCCCGCGGTGCGGATTCAGTCTGCGCGCCGACGTTGACGAATCTGTGACG	2548
DB	1410	AAGTACTGCTGCGCTGCGACGCGCGCACCTGCGCGCGCTGACGAATCTCTGACG	1351
QY	2549	ACCGAGTCAGAGCTTCGGTCCGGGGTGAACCGCGCGGAGCTAGCGGGTGGCCCCCGCC	2608
DB	1350	ACCTCGTGAAGGAGGTGTGCGGGGTGACGCGCGCGCGATGTAGCGGGTGGCGCGCTCC	1291
QY	2609	AGCCCGGGGAAACCGGCTTCCCGGTACAGGTAGACGTCTCCGAGAGAGTTCATCTGACAC	2668
DB	1290	AGGTGCGGAAAGCTCGGCTCGGGTACAGGTACAGTGTGCGAGGAGATTCACCTGTGACAC	1231
QY	2669	GCAACCTGAGATGGGCGGCTGCGGCGCATCTGTTCCGGCGGATCCGACACAGCTGGCG	2728
DB	1230	GCGACCTGCGGGTGGCGGTGGCGCGCATGTGTGGCGGCTTGATCCGACAGATTCGGCG	1171
QY	2729	TCGACACCCCGACGAGGCTCTCAACCGGTAAACCAAGTGTGTCTGCAATGCGCGGGTTC	2788
DB	1170	TCGGCGCCGAGTACGAGGCTGTTCAGGGCGGTAGCGGTGATGTGAGTCCGGGGGTG	1111
QY	2789	CGCTGCGGGCGGTATTCACGACCGGGGCGAGGCTTCCCGAACTCCGTCCGTTCGCC	2848
DB	1110	CGCTGCGGAGCCCGCTCTCTCGAAGGCGTTGAGGGCTCTCTGAGCTCGCGCTCTCC	1051
QY	2849	TCGACAACTGCGCGTCCGACCGGACCGCTGTAGTCTCTCGGACGCGTGAAGAGTGCAGC	2908
DB	1050	TCGGGAGCTTTCGCTGTCAAGCGCGCTGTAGTCTCTCGGAAATTTGAAGAGTGCATC	991
QY	2909	GCGCGGTGCGGCTGATCTGTTGAGTTTCGCGGATGAAGTTCACCAAGTGTGAAGCGCG	2968
DB	990	GTCCTGCGCTGCGCGCGCTGTGAGGTGCGGAGTGAAGTTCACCAAGTGTGAAGCGCG	931

QY 2969 TTGCGCGCGCGCGACAGATGATGTGTGAGCGCGACCGGATGCGCGCTTCG 3028
DB 930 GAGGCAACGCGCGCGGAGCAAGATGTAGGGGAACCGAGGTTGATTCGGCGATCTCGCGCTCG 871
QY 3029 GCGCGATTCGCAAGAAAGTCCCGGAGTCTTCTCTGACCGCTTCAAGAGGCGCACGTTG 3088
DB 870 GCGCGAGCTGTGAAGCGCGCGAGTCTTCGCGACGCGCGGAGAGGCGCTTCTCTG 811
QY 3089 CCGGTGTGCGCTGTGATCTGTGTGTGTGAGCGCGGACGAGGAGTGAAGACCGCGCG 3148
DB 810 CCGGTGTGCTGTGATCTCTGTGTGTGTGAGCGCGGACGAGGAGTGAAGTGTGCGTGC 751
QY 3149 AGGTCCACAGCGCGCGCTGCGCGCGAGCTGTGTGTGTGAGGAGGAGGAGGCTTGTG 3208
DB 750 AGCGCCGAGAGCGCGCGCTGCGCGCTCAAGGTGCGCTGCTGAGCGCGGAGGAGTGTG 691
QY 3209 TAGACGCTCAAGGCGAAGCGCGCGCGCGCGCGCTGCGCACCAAGCGCGCGGATACCGCG 3288
DB 690 TAGACGCTGCG 631
QY 3269 TTGCTCAACGCTTCAGAGCGCGCGCGGAGGATACATGCGCGCTTCTCGACGCGACG 3328
DB 630 TTGCTGAGCGGCTTCAGAGCGCGCGCGGAGGATACATGCGCGGAGGAGGAGTGTGCG 571
QY 3329 TCGTCAATGACCGAGGTGAGGATCTCGTTGCGCGGACCAACGACCTCGCGCTGACTTG 3388
DB 570 TCGTCAATGACCGAGCGGAAATGAGCGCTTCGCGCGCTGAGAGGCGAGCGGAGTGTG 511
QY 3389 GCACTGTTCACCGGTACCGGAAAGTGAACGCGGAAATGACATGCACTGCGCGCGCGGTAAAGC 3448
DB 510 GCGCGCTGTACAGCGAGCGAGAAAGTGCACCGGAAATGACATGCACTGCGCGCGCGGTAAAGC 451
QY 3449 CCGAGCGGTGACCGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3508
DB 450 CCGAGCGGTGACCGGAGGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
QY 3509 TTGACGCGCGCTGATGCTGTGTGCTTCAGATCTTTCGCGCGCGCGCGCGCGCGCGCG 3568
DB 390 TCGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 331
QY 3569 CCGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3628
DB 330 CCGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 271
QY 3629 TCGAACCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3688
DB 270 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 211
QY 3689 AGTTCGCGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3748
DB 210 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 151
QY 3749 GTGACTTCGCGGCTGAGAGCGCGGACCAAGCGCGCGCGGAGGAGGAGGAGGAGGAGGAG 3808
DB 150 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 91
QY 3809 TCGGAGATCTCGGTGAGTACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3859
DB 90 GGTTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 40

RESULT 12
US-10-160-539-24/c
; Sequence 24, Application US/10160539
; Publication No. US20030162262A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYPEPTIDE SYNTHASE
; FILE REFERENCE: 30062202120

PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 24
LENGTH: 1565
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-10-468-24

Query Match 8.2%; Score 729.4; DB 8; Length 1565;
Best Local Similarity 68.1%; Pred. No. 1e-153;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGGCGGTTCAGACGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2428
DB 1530 GGGGCGGTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1471
QY 2429 GCGACGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2488
DB 1470 GCGACGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1411
QY 2489 AAGAACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2548
DB 1410 AAGAACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1351
QY 2549 ACCGAGTCGACGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2608
DB 1350 ACCGAGTCGACGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1291
QY 2609 AGCCGCGGAGAACCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2668
DB 1290 AGCCGCGGAGAACCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1231
QY 2669 GCCACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2728
DB 1230 GCCACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1171
QY 2729 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2788
DB 1170 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1111
QY 2789 CGCTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2848
DB 1110 CGCTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 1051
QY 2849 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2908
DB 1050 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 991
QY 2909 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 2968
DB 990 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 931
QY 990 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 871
DB 2369 TTGCGCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3028
QY 930 GAGGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 871
DB 3029 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3088
QY 870 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 811
DB 3089 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3148
QY 810 GCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 751
DB 3149 AGGTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3208
QY 750 AGGTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 691
DB 3209 TAGAGGTTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3268
QY 690 TAGAGGTTCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 631

QY 3269 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3328
DB 630 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 571
QY 3329 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3388
DB 570 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 511
QY 3389 GCACTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3448
DB 510 GCACTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 451
QY 3449 CCGACGCTTCGACGCGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3508
DB 450 CCGACGCTTCGACGCGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 391
QY 3509 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3568
DB 390 TTGTCGACGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 331
QY 3569 CCGACGCTTCGACGCGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 3628
DB 330 CCGACGCTTCGACGCGCGCTTCAGACGCGCGCTTCAGACGCGATCAGCGCGATGCTCAACCCGTCG 271
QY 3629 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3688
DB 270 TCAGACGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 211
QY 3689 AGTTGCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3748
DB 210 AGTTGCGCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 151
QY 3749 GTGACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3808
DB 150 GTGACTGTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 91
QY 3809 TCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 3859
DB 90 TCGCGGTTCGCGCTTCAGACGCGCGCTTTCAGACGCGCGATCAGCGCGCGATGCTCAACCCGTCG 40

RESULT 14
US-10-846-335-24/c
Sequence 24, Application US/10846335
Publication No. US20050233431A1
GENERAL INFORMATION:
APPLICANT: ASHLEY Gary
APPLICANT: BERTLACH, Melanie C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002121
CURRENT FILING DATE: US/10/846,335
PRIOR APPLICATION NUMBER: US/04-05-14
PRIOR FILING DATE: US/09/793,708
PRIOR APPLICATION NUMBER: US 09/657,440
PRIOR FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: US 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/846,247
PRIOR FILING DATE: 1997-04-30
PRIOR APPLICATION NUMBER: US 60/134,990
PRIOR FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 24
LENGTH: 1565

TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-10-846-335-24

Query Match 8.2%; Score 729.4; DB 9; Length 1565;
Best Local Similarity 68.1%; Pred. No. 1e-153;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGGGGATGACGCGCTTCAAGACGCGGATCAAGCGCCCGGATGCTATCCACCCCTCG 2428
DB 1530 GGGGGGATGACGCGGCGGATGCGGATCAAGCGCGGATGCGGCGCTTCTCCACCGCTTC 1471
QY 2429 GCGACGTCCTCCGTTTCAAGTTGTTGAGCCGCGCGCTCAACCGATGTCGAAACCGTCGAGG 2488
DB 1470 GCGGCGTCGCGCTCAAGCTGTTTCAAGCGCGGCGGATGACGATGATGAAAGCGCTCAAG 1411
QY 2489 AAGAACTGTCCTCCCGCGCTGCGATGATGTCGCGCGCGGATGACGAAAGTCGTGACG 2548
DB 1410 AAGTAATCGTCGCGCTGACGCGCGCGGATGTCGCGCGCGCTGACGAAAGTCGTGACG 1351
QY 2549 ACCGAGTCGACGCGCTGCGGTCGCGGTCGACCGCGCGCGGATGACGCGCGCGCGCGCG 2608
DB 1350 ACCTCGATGAGGAGGATGTCGCGGATGACGCGCGCGGATGACGCGCGCGCGCGCGCG 1291
QY 2609 AGCCCGGAGGAAACCGGCTTCCGATCAAGTTAGACGTCGTCGACGAGTCGATCTGACG 2668
DB 1290 AGGTCGCGGAGGAGCGGCTTCCGATCAAGTTAGACGTCGTCGACGAGTCGATCTGACG 1231
QY 2669 GCCACTGCGGATGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 2728
DB 1230 GCGACCTGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 1171
QY 2729 TCGACACCCCGGAGGAGCTTCCACAGCTTACAGGTCGTCGTCGTCGTCGTCGTCGTC 2788
DB 1170 TCGGCTCCGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 1111
QY 2789 CGCTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2848
DB 1110 CGCTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1051
QY 2849 TCGGACACCTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2908
DB 1050 TCGGACACCTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 991
QY 2909 GCGCGTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2968
DB 990 GTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 931
QY 2969 TTGCGCGCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3028
DB 930 GAGGACACGCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 871
QY 3029 GCGCGATCTCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3088
DB 870 GCGCGATCTCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 811
QY 3089 CCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3148
DB 810 CCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 751
QY 3149 AAGTCCACAGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3208
DB 750 AAGTCCACAGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 691
QY 3209 TGAACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3268
DB 690 TGAACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 631
QY 3269 TTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3328
DB 630 TTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 571
QY 3329 TCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3388

DB 570 TCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 511
QY 3389 GACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3448
DB 510 GCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 451
QY 3449 CCGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3508
DB 450 CCGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 391
QY 3509 TTGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3568
DB 390 TCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 331
QY 3569 CCGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3628
DB 330 CCGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 271
QY 3629 TCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3688
DB 270 CCGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 211
QY 3689 AGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3748
DB 210 AGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 151
QY 3749 GTGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3808
DB 150 GCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 91
QY 3809 TCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3859
DB 90 CCGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 40

RESULT 15

US-09-861-289-9/c
Sequence 9, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 9
LENGTH: 1458
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-9

Query Match 8.1%; Score 722; DB 3; Length 1458;
Best Local Similarity 69.5%; Pred. No. 4.7e-152;
Matches 980; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 2395 GCGATACAGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2454
DB 1455 GCGACGAGAGCGCGGCTCTTCCAGCGCTTCCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 1396
QY 2455 CCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2514
DB 1395 GCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1336
QY 2515 GATGTCGCGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2574

Db 1335 CACCTGCGCGCGCTGCAAGATCCTGAGCAAGCTTGAGAGGATGTCGAGGAGT 1276
Qy 2575 GACCCGCGCGCGCAATGATGGGGTGGCCCCCGCCAGCCCGGGGAAACCGGCTTCCGGTA 2634
Db 1275 CACGCGCGCGCGCAATGATGGGGTGGCCCCCGCCAGCTCGAGGAAAGCCGCTCGCGTA 1216
Qy 2635 CAGGTAGACGCTCGAGCAGATGATCTGCAACCGCACCTTGCGGATGGGCGGTGCGGCG 2694
Db 1215 CAGGTAGACGCTCGAGCAGATGATCTGCAACCGCACCTTGCGGATGGGCGGTGCGGCG 1156
Qy 2695 CATCGTTTCGCGCGGATCCGCAACAGCTGCGGTGCAACCCCGACGAGGCTCTCA 2754
Db 1155 CATGCTGCGCGCTTGATCCGACAGATCGGCTGCGCCCGGCTGCGAGGCTGTTCA 1096
Qy 2755 CGCGTACCCAGATGCTGCTGCAATGCCCGGAGTCCGCTCGCGCGCTGATGCAAGAACG 2814
Db 1095 GCGGTACCGGTAGTGTGATGTGAGATCCGAGGAGTCCGCTCGCGAGCCGCTCTGAAAGCC 1036
Qy 2815 GCGAGGCGCGCTCCGCAACTCGCTCGCTCGGACCAACCTGCGCTGTCGCAAC 2874
Db 1035 GTTGAAGGCGCTCTGAGAGCTCGGCGCTCTCTCTGCGCGAGCTTGCGCTGTCACGCGC 976
Qy 2875 GCTGTAGTCTCTGCGGAGCGGTGCAAGTCAAGTCAAGCGCGCGGTGCGGCTGCACTGTTGAG 2934
Db 975 GCTGTAGTCTCTGCGGAAATGTTGACAAATCGATGCTGCTGCTGCGCGGTGTTGAG 916
Qy 2935 TTGCGCGATGAAATGTCACAGATGATGAGCGGTTGCGCCCGCGCAGAGATGATG 2994
Db 915 GTTCGCGCAATGAAATGTCACAGATGATGAGCGGTTGCGCCCGCGCAGAGATGATG 856
Qy 2995 GTTGAAGCGCGAGCGAGCCGCTGCGCGCGCTTCCGCGCGCATCCGCAAGATCCCGAG 3054
Db 855 GCGCAAGCGCGAGGTTGATCGCGCATCGCGCTTCCGCGCGCATGCTGAGAGCGCGCAG 796
Qy 3055 GTTCTTCCTGACCCGTTGAAAGCGCGCACGCTTGCGGTTGAGCTGTACTGTCGTT 3114
Db 795 GTTCTTCGCGAGCGCGCGGAAAGCGCGCTTCTTGCGGTTGTTGCTGTACTCTGTC 736
Qy 3115 GTTCAAGCCCGTACAGAGAGTACGACCGCGCGAGGTCCTCAAGCCCGGCTGCGCG 3174
Db 735 GTTGAAGCGCGTGAAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 676
Qy 3175 CAGCGTCTGTTGCGGTGAAGGCGAAAGCGTTGTTGATGACGCTGCAAGGCGAAAGCCCGCG 3234
Db 675 CAGGTTGCTGCTGCTGAGCGCGAAAGATTCGTGTAGACGCTGAGCGCGCAAGCCGTGCTC 616
Qy 3235 GCGCGCGTGAACAACAAGCCCGCATACCGGTTGTTGATCAAGGTTCCAGGCGCGCGA 3294
Db 615 GGTGGCGTGGCGCGCAGGCTCCGAGGCGCGGAGTTGTTGAGCGGCTCAAGCGCGCGA 556
Qy 3295 GAGGTACATCGCCCTGCGGTTCTCGACGCGCATCTGTCGATGACCGGAGTGAAGATCTC 3354
Db 555 GAAGTACATCGCCCGAGGAGTTGCGCGGATATCTGTCGATGACCGGACCGGAAATATGCGC 496
Qy 3355 GTTGGCGGAGCAACAAGCTCGCGCTGTAATTGACCTGTCACCCGTAACGAGAAATG 3414
Db 495 GTTGGCGGAGTGAAGGCGAGAGGTCGTAGCGGCGCGGTCACACGAGACGAGAAATG 436
Qy 3415 ACAAAGCAATGCAAGTGGCGCGGAGTGAAGCCGTAACGCTGTACGAGAAAGCGGCTC 3474
Db 435 GCAAGGCAATGCAAGTGGCGCGGAGTGAAGCCGTAACGCTGTACGAGAAAGCGGCTT 376
Qy 3475 CCGCGTACCGAGCTTCAAGCGCGCGCGCGCGGCTGTCAGCGGCTGATCTGTTGCTCA 3534
Db 375 CCGCGAGCGCGCGGTGAAAGACCGCGCTGTTGAGCGGAGCAGGAGTTCCTTCA 316
Qy 3535 GTACTTGGCCCGCGGCGCTGTCACCGCGCGTGCAGCTGCGGAGTGCATGCAACAG 3594
Db 315 GTACGCGCGCGCGGCGGTTCGACCGCGGTGCGAGCTTCGAGAACTTGCCGAAACAG 256
Qy 3595 GTGAGCAGCTTCGCGAAACCGCGCGCGGTTCAGGTGAAACGACCGCGCATCTGCTCA 3654
Db 255 GCGAGGAGGCGCGGAAAGCGTCCCGGTGACGCGCAGGTGTGCGGCGCTCTCCAG 196

Qy 3655 CGGAGTGAACGCGGAGGCGCGGTAGTGGCGGTGATTCGCGAGCGAGCCGCTGTCG 3714
Db 195 CGGAGTGAACGCGGCTTGTGCTGATGCGCACGCGAGACGAGGTTGCGGCGGTGCT 136
Qy 3715 TTGCGCGGTGTCGATGCGGACCGGTGAGGCGGAGTGAACCTCGGCTGTAACCGCGCAC 3774
Db 135 TCCGCGCTGTCGAGGCGGACAGAGCGCGCGCGGAGGATCTGCGCGACGCGGTGAC 76
Qy 3775 CAGGCGCGCGGTGCGGCGGTGCGGCTGCGGCTGCG 3804
Db 75 CGCGCGCGCAATCGGCTCGGAGTGC 46

Search completed: April 7, 2006, 04:36:08
Job time : 6273.93 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:28:59 ; Search time 1488.43 Seconds
(without alignments)
10776.903 Million cell updates/sec

Title: US-10-611-442-1
Perfect score: 8942
Sequence: 1 ggcgccttcgaccatcgcg.....gtggagaccggagagatcc 9024

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/1 COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5 COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/H COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq: *
7: /cgn2_6/ptodata/1/ina/PP COMB.seq: *
8: /cgn2_6/ptodata/1/ina/RE COMB.seq: *
9: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	3032.6	33.9	8051	2	US-08-576-626A-2
C 2	2606.8	29.2	47981	3	US-09-679-279-1
C 3	1309.4	14.6	13613	3	US-09-105-537-3
C 4	729.4	8.2	1565	3	US-09-320-878-24
C 5	729.4	8.2	1565	3	US-09-141-908-22
C 6	729.4	8.2	1565	3	US-09-657-440-24
C 7	729.4	8.2	1565	3	US-09-793-708-24
C 8	722	8.1	1458	3	US-09-105-537-9
C 9	701.4	7.8	3756	2	US-08-576-626A-1
C 10	595.4	6.7	1248	3	US-09-105-537-7
C 11	519.2	5.8	5970	3	US-09-320-878-21
C 12	519.2	5.8	5970	3	US-09-141-908-11
C 13	519.2	5.8	5970	3	US-09-657-440-21
C 14	519.2	5.8	5970	3	US-09-793-708-21
C 15	461.2	5.2	1452	3	US-09-758-759-30
C 16	461.2	5.2	109519	3	US-09-758-759-1
C 17	397.4	4.4	80161	3	US-09-036-987A-1
C 18	397.4	4.4	80161	3	US-09-370-700-1
C 19	397.4	4.4	80161	3	US-09-603-207-1
C 20	381.6	4.3	6085	3	US-09-029-603-4
C 21	371	4.1	1518	2	US-08-660-765A-1
C 22	350.2	3.9	4496	3	US-08-765-907A-6
C 23	350.2	3.9	4496	3	US-09-987-614A-6
C 24	346.2	3.9	1194	3	US-08-765-907A-9

25	346.2	3.9	1194	3	US-09-987-614A-9
26	266.2	3.0	50917	3	US-09-428-517-1
27	255	2.9	11440	3	US-09-252-991A-2912
C 28	255	2.9	1425	3	US-09-252-991A-3053
29	238.2	2.7	38506	3	US-09-320-878-19
C 30	238.2	2.7	38506	3	US-09-141-908-1
31	238.2	2.7	38506	3	US-09-657-440-19
C 32	238.2	2.7	38506	3	US-09-793-708-19
C 33	212.6	2.4	1600	3	US-09-434-288-10
C 34	207.8	2.3	1281	3	US-09-105-537-19
C 35	196.4	2.2	8051	3	US-08-576-626A-2
C 36	192.8	2.2	12249	3	US-09-266-965-74
C 37	192.8	2.2	18331	3	US-09-266-965-96
C 38	189.6	2.1	44377	2	US-08-804-227C-7
C 39	189.6	2.1	44377	2	US-08-804-198-1
C 40	187	2.1	1224	3	US-09-266-965-22
C 41	183.2	2.0	1188	3	US-09-252-991A-936
C 42	182.8	2.0	1332	2	US-08-660-765A-3
C 43	182.4	2.0	588	3	US-09-724-797-45
C 44	180.8	2.0	47981	3	US-09-679-279-1
C 45	178.6	2.0	1209	3	US-09-105-537-21

ALIGNMENTS

Sequence 9, Appl1
Sequence 1, Appl1
Sequence 2912, Ap
Sequence 3053, Ap
Sequence 19, Appl1
Sequence 1, Appl1
Sequence 19, Appl1
Sequence 19, Appl1
Sequence 10, Appl1
Sequence 19, Appl1
Sequence 19, Appl1
Sequence 74, Appl1
Sequence 96, Appl1
Sequence 7, Appl1
Sequence 1, Appl1
Sequence 22, Appl1
Sequence 936, App
Sequence 3, Appl1
Sequence 45, Appl1
Sequence 1, Appl1
Sequence 21, Appl1

RESULT 1
US-08-576-626A-2/c
Sequence 2, Application US/08576626A
Patent No. 5928194
GENERAL INFORMATION:
APPLICANT: Summers, R.G.
APPLICANT: Katz, U.
APPLICANT: Donadio, S.
APPLICANT: Staver, M.J.
TITLE OF INVENTION: POLYMERIDE-ASSOCIATED SUGAR
TITLE OF INVENTION: BIOSYNTHESIS GENES
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSER: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/576, 626A
FILING DATE: 21-DEC-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Diane Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 5857.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 938-3137
TELEFAX: (847) 938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8051 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-576-626A-2

[illegible]

Db	6484	CGCCGACATGTAATGCTTGGAGTTGTCCTGGGACACTTGTCATGATGCGGCCA	6425
Oy	3348	GGATCTGTTGCCCGGACCAACGCACTCCGCGTGTACTTGGACCTGTACCCGTACGC	3407
Db	6424	GCGTCTGTTGCCCGGACCGAATCGGCGCTGTAGCGGACACCGGTACCCGACGC	6365
Oy	3408	AGAGTGAACAGCGGAACATGCACTCGGCGCCGGGTTAGAGCCGACGCTGTACGGAAAG	3467
Db	6364	AGAGTGAACAGCGGAACATGCACTCGGCGCCGGGTTACAGGCCACGCTGTAGGGAAAG	6305
Oy	3468	CGGCTCCCGCGGTAACCGGACGCTTCAACGCGCGGCGGCGTTAGGCGGCTGATCGTGT	3527
Db	6304	CAGGCTTGGGTAGACCGCGCGCTCCAGTGGACCTTGGCGGTCAGCGGCTTATCGTGT	6245
Oy	3528	TGCTTCAGACTTTCGCCCGCGGCGCTTGTCCACCGCGGTGCCCAGCTTGGGATGCGAT	3587
Db	6244	TGACACAGTACTTTCGCCCGCGGACCGTGTCCACCGCGGTGCCAGGTCCGGATGCGGC	6185
Oy	3588	CGAACAGTGCAGCAGCTCGCCGAACCCCGGCGCGGTCCAGGTGCAACCGAGCGCATCT	3647
Db	6184	CGAACAGTGCAGCAGGTGGCGGAATCCGCGCGGTCCAGGCCACGCGCGCACGCTCT	6125
Oy	3648	GCTCAACCGGGGTAAACGCGGAGCGCGGTAGTGGGCGGTGATTCGGCGGACCGGACG	3707
Db	6124	GCTCAAGCGGTAAACGAGTGGCGCCCGCTGTATGCGGCGAGGTCCGCGACCGGCGG	6065
Oy	3708	CCTGTGCTTCGCGCGGTGTCGATCGGACCGGACCGGTGAGCGCGGTGACCTCGCGCTGAGCG	3767
Db	6064	CTTCGACGTGGGCGTCAACCGTCAACCGGCGCGGTGACCGCGGAACTCCGAGCTTGACG	6005
Oy	3768	CCCGCACCAACGCGCGCGCGGTCCGCGGTGGGTGTGTCGCGCGCTCCGCACTTCGTCGCTA	3827
Db	6004	CCCGCACCAACGCGCGCGGTCCGCGGTGACGTCGCGCGCGCGCTCCGCAACCCCGCTTCT	5945
Oy	3828	CGGCGGTCCCGGTAGAGTGGTCTTCAATCGACGTGCGAAACCTTCTGAGCGTGTGGTGC	3887
Db	5944	GCGCGGTTCGCTACGAGTGTGTTCATCGGGAGCGAACTCTGGCACCTTGTGGCGCA	5885
Oy	3888	GAGGATCAACGAACGTTGCGTTTCGCGTGTCCCACTCCGCGTGTATCAACGACCGCTG	3947
Db	5884	GGTGTCACGAACGTTTGGCGGTCTCTGTCCTCATTTGCGGTTGATCAGCTCCGCTG	5825
Oy	3948	GTGTGTCGAGTCGAGTACGTTCGACACCCGCGGATGTCTTCACTGAGAACCGAAGGG	4007
Db	5824	GTGGCGGACCGCGATGATGTCCACACCCGCGGATGTCTTGTGAGAACCGCGGGG	5765
Oy	4008	CCGTCGCGGAAGGAGACACCGTTTCCGCGGACCACTCGGTCTGTGACCGGACGCGT	4067
Db	5764	CCGTCGCGCAACGCGACGACCCCGGTCCGAGAGCTTGTGGGTCTCCAGCGCGACGCGC	5705
Oy	4068	GAGTCGAGCGGTAGGCGTCACTGTGTGCGAGCGCGGAGGAGAGTAGAGGCTGTGCAACG	4127
Db	5704	GAGTCGAGTCGCGTACGCTGATCTGTGTGACACCCGCGGAGAGAGTAGAGGTTGTGCA	5645
Oy	4128	ACCTTCTTCGCGCGCAGGATGCTCTGACCCGCTGACCGTGTACGTCGATGCGCGTGGCGG	4187
Db	5644	ACCTTCTTCGCGCGCAGGATGCGCTGCAACTGTGTGCGGTGAGTGCCTGTGCGCGGAG	5585
Oy	4188	TCCACCAAGATGATCCAGTACTCGTAGTTGTTCTGTCTGCGCAAGATCCACGCGTGCAG	4247
Db	5584	TCCACCAAGATGATCCAGTACTCGTAGTTGTTCTGTCTGCGCAAGATCCACGCGTGCAG	5525
Oy	4248	GTGACGCGCGTACGTCGCGACGCTCGCTGTGTAGACGCGGTGTTAGCCCGTTGTGC	4307
Db	5524	GATATGCGCGCACGTCCTCCGAGACCTCGTCGAGTAGAGCGGTGTGAGCGGTTGTGCG	5465
Oy	4308	TCCCTGATCTCGCGCAACGCGGTGAGGGAGGTAGACCCCATGGCCGCGCGCACTCGTTC	4367
Db	5464	ACCTTGATCTCGCGCAACGCGGTGAGCGGATGAGGCCATTCGCGCGCGCGCACTCGTTC	5405
Oy	4368	ATCTTGGCTGTGTGTCGATCTCGGTACCACTTGTCCGGGCGGATGCCAAGTTGTGC	4427

Db 5404 ATCTTCCGCTGTGTCAGACATCGATCAAGCTTGTCCGCTGCGATCCGGAAGTTGTC 5345
 Qy 4428 ATGACCCGGAATCCGTTGCGCAGACGCGCTGCTGCTGATCCACCCGCCCCCTTCGAG 4487
 Db 5344 ATGACCGGAATCCGTTGCGCAGACGCGCTGCTGCTGATCCACCCGCCCCCTTCGAG 5285
 Qy 4488 GCGGTGACCGCTTGTGTGCGGTGGAAGCTGAACACTTCGCGCTGACCGGATCCCGCAC 4547
 Db 5284 GAGGTGACCGCTTGTGTGCGGTGGAAGCTGAACACTTCGCGCTGACCGGATCCCGCAC 5225
 Qy 4548 GGAAGTCAACCCGCTGTCGACGCCAGGCGCTGCGCGCTGGAAGAAAGACTTGAACCTGG 4607
 Db 5224 GCGCGCCGCGCGGTGTCAGCCAGCGCGCTGCGCGCTGGAAGAAAGACTTGAACCTGG 5165
 Qy 4608 TGGTCCGCGCGGAATCTTCCGACGCGCTGCAAGGTCTGTGCGGCCCCCAAGTTGACG 4667
 Db 5164 TGGTCCGCGCGGAATCTTCCGACGCGCTGCAAGCGCGCGCTGCGCCCCAGAGTTGACG 5105
 Qy 4668 CCGACGATCCGCGCGCTGTCGCGGCTGACGACGCGCGCTGACGCTGCTGCGGCTGACG 4727
 Db 5104 CCGATGATCCGCGCGCTGTCGCGGCTGACGACGCGCGCTGACGCTGCTGCGGCTGACG 5045
 Qy 4728 CCGGTGCGCGGCTGACGCTGCGAGAACCGGTGTGATCCAGCCAGCTTCGCGCGTGC 4787
 Db 5044 CCGGTCTCGGCGCTGACGCTGCGAGAACCGGTGTGATCCAGCCAGCTTCGCGCGTGC 4985
 Qy 4788 GCGGTGCGCGGCTGACGCTGCGAGAACCGGTGTGATCCAGCTTCGCGCGTGC 4847
 Db 4984 GCGGTGCGCGGCTGACGCTGCGAGAACCGGTGTGATCCAGCTTCGCGCGTGC 4925
 Qy 4848 AGCACGATTCAGAGGCGACGCTGCGAGTTCAGGTGCGATTCGAGTCCGTAACCCGAC 4907
 Db 4924 AGCACGATTCAGAGGCGACGCTGCGAGTTCAGGTGCGATTCGAGTCCGTAACCCGAC 4865
 Qy 4908 AGGTGCGCGACCGCGGCTGGAATCCCGACACAGGCGCGCGCTGCTGAGGCACTTG 4967
 Db 4864 AGGTGCGCGACCGCGGCTGGAATCCCGACACAGGCGCGCGCTGCTGAGGCACTTG 4805
 Qy 4968 TGTTCAGAGGCGCACTTCAGGCGGCGAGAACCGCTGCGCTGCTGAGTCCGTCGCGG 5027
 Db 4804 TGTTCAGAGGCGCACTTCAGGCGGCGAGAACCGCTGCGCTGCGTCGCGTCGCGCTG 4745
 Qy 5028 CCGACGTCAGAGGCGGTCAGAGAACGCTTCGAGACCGCGAGAACGTCAGATCGATCGGT 5087
 Db 4744 CCGACGTCAGAGGCGGTCAGAGAACGCTTCGAGACCGCGAGAACGTCAGATCGATCGGT 4685
 Qy 5088 AGCGGCTTCAAGCGCTGCGCGCTGCTGAGACCGCGAGACCGCGAGACCGAGCTGCGCA 5147
 Db 4684 GCGCGTTTCATCCGCGCGCTGCTGCTGAGACCGCGAGACCGCGAGACCGAGTCTGCGCA 4625
 Qy 5148 GCTGATGTTGATGATGCTGCTGCGCGAGACGCTTCGATCTGCGAGTCCGATCC 5207
 Db 4624 CCGTGAATGTTGATGATGCTGCTGCGCGAGACGCTTCGATCTGCGAGTCCGATCC 4565
 Qy 5208 ACCGCTGTCGAGGCGACCGCGCTGCTGAGAACGCGCGGCGACCTGAGAGATGATAC 5267
 Db 4564 ACCGCTGTCGAGGCGACCGCGCTGCTGAGAACGCGCGGCGACCTGAGAGATGATAC 4505
 Qy 5268 GGTTCCTGTTGCGGTGAGAACCGCGCAACCTTCCTGAGAGTCCAGGCTGCTGAGATCT 5327
 Db 4504 GGTTCCTGTTGCGGTGAGAACCGCGCAACCTTCCTGAGAGTCCAGGCTGCTGAGATCT 4445
 Qy 5328 GGTTCAGCGGAGACGTCAGACGATGTCAGAGTTCGAGGCTGCGGCGGTCTTCGCGGACGT 5387
 Db 4444 GCGAGGCGCGGACGTTCAACATGATTCACATAGGCGCGGTTCGGAACCTTCGCGGCGGT 4385
 Qy 5388 CCGTGTAGTTGTCGCGCTGACGATGACGCTGCGCGCACTGCGGAGACCTTCGCGGAC 5447
 Db 4384 CCGCTGTAGTTGTCGCGCTGACGATGACGCTGCGCGCACTGCGGAGACCTTCGCGGAC 4325
 Qy 5448 CCGCTGTGTCGAGGTCAGACGAGCGCTGAGAGGCTGCTGATCTCTTGAACAGCA 5507
 Db 4324 CCGCTGTGTCGAGGTCAGACGAGCGCTGAGAGGCTGCTGATCTCTTGAACAGCA 4265

Qy 5508 GGGGAGACATGCGCTGCTGCGGCGGAGAGACCGGTTGATTCAGAGACCTCACTTCG 5567
 Db 4264 GGGGAGACATGCGCTGCTGCGGCGGAGAGACCGGTTGATTCAGAGACCTCACTTCG 4205
 Qy 5568 GGTGACTGTGCTGTCAGAGACCGCGAGAGTCTGTAAGTCTTCTGCTGCTGCTGCA 5627
 Db 4204 GGTGCTGTGCTGTCAGAGACCGCGAGAGTCTGTAAGTCTTCTGCTGCTGCTGCA 4145
 Qy 5628 TGGCTGTGCTGCGCGGATTCACCCGCTGCGTGTGATGTCGACCAAGCGCGGTGCTT 5687
 Db 4144 TGGCTGTGCTGCGCGGATTCACCCGCTGCGTGTGATGTCGACCAAGCGCGGTGCTT 4085
 Qy 5688 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 5747
 Db 4084 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 4025
 Qy 5748 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 5807
 Db 4024 TGGCTGTGCTGCGCGGATTCACCCGCTGCGTGTGATGTCGACCAAGCGCGGTGCTT 3965
 Qy 5808 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 5867
 Db 3964 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 3905
 Qy 5868 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 5927
 Db 3904 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 3845
 Qy 5928 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 5987
 Db 3844 GGAAGAACCTCAAGTGTGCTGCTGATGTCGAGATTCAGAGTCCAGAGGCTTCAATGTTGT 3785
 Qy 5988 TGTTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6047
 Db 3784 TGTTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3725
 Qy 6048 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6107
 Db 3724 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3665
 Qy 6108 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6167
 Db 3664 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3605
 Qy 6168 TGTTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6227
 Db 3604 TGTTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3545
 Qy 6228 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6287
 Db 3544 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3485
 Qy 6288 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6347
 Db 3484 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3425
 Qy 6348 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6407
 Db 3424 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3365
 Qy 6408 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6467
 Db 3364 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3305
 Qy 6468 TGGCTGTGCTGCGCGGATTCACCCGCTGCGTGTGATGTCGACCAAGCGCGGTGCTT 6527
 Db 3304 TGGCTGTGCTGCGCGGATTCACCCGCTGCGTGTGATGTCGACCAAGCGCGGTGCTT 3245
 Qy 6528 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 6587
 Db 3244 GATTCGAGTTCGCTTTCGAGAGAACCGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 3185

```
OY 6588 CCGCGGAGTACCGCCGCGAGTGCATTCGATCAGACCC-CGCCAGGTAGAGGCTCC 6646
Db 3184 CCGGTTGCTGCGCCGCGGCG-----GTGTGATCAGACCGCGAGCTACTGTAAAGGCTCC 3130
OY 6647 GCCCGCAGCGGAGAGAGCTCC 6667
Db 3129 GGTCAACGCGCGGAGAAAGGCC 3109

RESULT 2
US-09-679-279-1
; Sequence 1, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Voichegursky, Yanna
; TITLE OF INVENTION: Recombinant Megalomycin Biosynthetic
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 47981
; TYPE: DNA
; ORGANISM: Micromonospora megalomicea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (144)
; OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
; OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (926) ... (2061)
; OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-1somerase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-1somerase;
; OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (2072) ... (3382)
; OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
; OTHER INFORMATION: TDP-megosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (3462) ... (4634)
; OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase
; OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (4651) ... (5775)
; OTHER INFORMATION: megDII, deoxybugar transferase (eryCI, DmrJ homolog),
; OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-amino transferase;
; OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (5822) ... (6595)
; OTHER INFORMATION: megDII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
; OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (6592) ... (7197)
; OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmV hc
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
; OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (7220) ... (8206)
; OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (8228) ... (9220)
; OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
; OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
;
; NAME/KEY: CDS
; LOCATION: (9226) ... (10479)
; OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (10483) ... (11424)
; OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (12181) ... (22821)
; OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (12505) ... (13470)
; OTHER INFORMATION: megAI, AT-L
; NAME/KEY: misc_feature
; LOCATION: (13576) ... (13791)
; OTHER INFORMATION: megAI, ACP-L
; NAME/KEY: misc_feature
; LOCATION: (13849) ... (15126)
; OTHER INFORMATION: megAI, KS1
; NAME/KEY: misc_feature
; LOCATION: (15427) ... (16476)
; OTHER INFORMATION: megAI, AT1
; NAME/KEY: misc_feature
; LOCATION: (17155) ... (17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc_feature
; LOCATION: (17847) ... (18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc_feature
; LOCATION: (18268) ... (19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc_feature
; LOCATION: (19876) ... (20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc_feature
; LOCATION: (21517) ... (22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc_feature
; LOCATION: (22318) ... (22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867) ... (33555)
; OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (32957) ... (42437)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc_feature
; LOCATION: (24544) ... (25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc_feature
; LOCATION: (26230) ... (26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc_feature
; LOCATION: (26998) ... (27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc_feature
; LOCATION: (27353) ... (28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc_feature
; LOCATION: (28897) ... (29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc_feature
; LOCATION: (29953) ... (30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc_feature
; LOCATION: (31396) ... (32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc_feature
; LOCATION: (32257) ... (32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc_feature
```

```

1 LOCATION: (33052)...(33312)
2 OTHER INFORMATION: megAIII, ACP4
3 NAME/KEY: CDS
4 LOCATION: (33666)...(43271)
5 OTHER INFORMATION: megAIII; SEQ ID NO: 15= translated amino acid sequence
6 NAME/KEY: misc feature
7 LOCATION: (33760)...(35027)
8 OTHER INFORMATION: megAIII, KSS
9 NAME/KEY: misc feature
10 LOCATION: (35385)...(36419)
11 OTHER INFORMATION: megAIII, ATS
12 NAME/KEY: misc feature
13 LOCATION: (37068)...(37604)
14 OTHER INFORMATION: megAIII, KRS
15 NAME/KEY: misc feature
16 LOCATION: (37860)...(38120)
17 OTHER INFORMATION: megAIII, ACP5
18 NAME/KEY: misc feature
19 LOCATION: (38187)...(39470)
20 OTHER INFORMATION: megAIII, KSG
21 NAME/KEY: misc feature
22 LOCATION: (39795)...(40811)
23 OTHER INFORMATION: megAIII, AT6
24 NAME/KEY: misc feature
25 LOCATION: (41406)...(41936)
26 OTHER INFORMATION: megAIII, KR6
27 NAME/KEY: misc feature
28 LOCATION: (42168)...(42425)
29 OTHER INFORMATION: megAIII, ACP6
30 NAME/KEY: misc feature
31 LOCATION: (42585)...(43271)
32 OTHER INFORMATION: megAIII, TE
33 NAME/KEY: CDS
34 LOCATION: (43268)...(44344)
35 OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
36 OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
37 NAME/KEY: CDS
38 LOCATION: (44355)...(45623)
39 OTHER INFORMATION: megCII, desosamine glycosyltransferase
40 OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
41 NAME/KEY: CDS
42 LOCATION: (45520)...(46591)
43 OTHER INFORMATION: megBII-2 (megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
44 OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
45 OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
46 NAME/KEY: CDS
47 LOCATION: (46660)...(47403)
48 OTHER INFORMATION: megH, TRII; SEQ ID NO: 19= translated amino acid sequence
49 NAME/KEY: CDS
50 LOCATION: (47411)...(47980)
51 OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888

```

QY	6655	GGGAGAAAGTTCCTTCGAACAACCTTCCGGGTGACCCGGTCCCGCGCTCCGTGTGAACACGGGC	6714
Db	241	CGGAGAAAGTTCCTTCGAACAACCTTCCGGGTGACCCGGTCCCGCGCTCCGTGTGAACACGGGC	300
QY	6715	GTCCGAGACCCCGATCAATGCTGTCCGGTGAACCTTCTTAACGTGTGGGCGCGACATCTTTTC	6774
Db	301	GTCCGAGACCCCGATCAATGCTGTCCGGTGAACCTTCTTAACGTGTGGGCGCGACATCTTTTC	360
QY	6775	TGACCCGATGTTCCTGTGTATGAACGCGCTTCCCGCCCGCTCTGTGAACCTGTGCTGTGAGACT	6834
Db	361	TGACCCGATGTTCCTGTGTATGAACGCGCTTCCCGCCCGCTCTGTGAACCTGTGCTGTGAGACT	420
QY	6835	GACCGGTTGCGCGGCTGTTTTCGCCCTTTCGAACTGCGGATTTGCTGATCGCGCAAGTG	6894
Db	421	GACCGGTTGCGCGGCTGTTTTCGCCCTTTCGAACTGCGGATTTGCTGATCGCGCAAGTG	480
QY	6895	GGAGCGGGGTGCTGAACCGGGATGATCTGTGCAATCATGCGGCTCAATGACATCTCTTATAG	6954
Db	481	GGAGCGGGGTGCTGAACCGGGATGATCTGTGCAATCATGCGGCTCAATGACATCTCTTATAG	540
QY	6955	CATGATCCGCGCGAGGGATCCGACAGGCCCGAAACGCGCGACATCAAGCTGTGTGACGA	7014
Db	541	CATGATCCGCGCGAGGGATCCGACAGGCCCGAAACGCGCGACATCAAGCTGTGTGACGA	600
QY	7015	CGTGCAATCAATCCGTGCAAGCGCGGATGACACCGACACACAGCCATCTGTGTCGCACT	7074
Db	601	CGTGCAATCAATCCGTGCAAGCGCGGATGACACCGACACACAGCCATCTGTGTCGCACT	660
QY	7075	GGAAAGGGTGGCGGATCAGGGGAAATGGCCCGTGTCACTAGACAGACGCGCAACAGCTGTCC	7134
Db	661	GGAAAGGGTGGCGGATCAGGGGAAATGGCCCGTGTCACTAGACAGACGCGCAACAGCTGTCC	720
QY	7135	GGGCTTCGCGAAACAGCATGATCTGTGTGACGCGCTTCAATTGCCCGCGCGACCGGCTT	7194
Db	721	GGGCTTCGCGAAACAGCATGATCTGTGTGACGCGCTTCAATTGCCCGCGCGACCGGCTT	780
QY	7195	GGAAATCCGTGCCACCGGTCGTCCGACGTACGATCGCGGACCCGGGTTTTCAGACAGCA	7254
Db	781	GGAAATCCGTGCCACCGGTCGTCCGACGTACGATCGCGGACCCGGGTTTTCAGACAGCA	840
QY	7255	GGTATGAGGGGATGACAGCGCTTTCGTCTGCGCGCGACCGGTGGCACTTACGTGGAATCCG	7314
Db	841	GGTATGAGGGGATGACAGCGCTTTCGTCTGCGCGCGACCGGTGGCACTTACGTGGAATCCG	900
QY	7315	TCAACAGTCTTCAATCCCGGAGCGTTCCTAATGGCAATTGGACATCGAAGCGGACTGGGCGG	7374
Db	901	TCAACAGTCTTCAATCCCGGAGCGTTCCTAATGGCAATTGGACATCGAAGCGGACTGGGCGG	960
QY	7375	GAGTTGACAGATGGCCCGGGGTCTTCTACTGGGAGGTTCCGTGCCAAACCGCGATCTGTACTCG	7434
Db	961	GAGTTGACAGATGGCCCGGGGTCTTCTACTGGGAGGTTCCGTGCCAAACCGCGATCTGTACTCG	1020
QY	7435	ATGCTCTGTTCGGAACGGGACGACGACCCCTGTGACTGTGTACGAACGGTTGCGGCGCC	7494
Db	1021	ATGCTCTGTTCGGAACGGGACGACGACCCCTGTGACTGTGTGTACGAACGGTTGCGGCGCC	1080
QY	7495	GGAACGGGACCGTATACGCACTGCGGCGCGAAACGTGGGTGCTGAGTGAACACCGGACCGCC	7554
Db	1081	GGAACGGGACCGTATACGCACTGCGGCGCGAAACGTGGGTGCTGAGTGAACACCGGACCGCC	1140
QY	7555	GCCGAGGTGTGCGCCGATCCGAGGCTTCAACCAAGGCGCGCGCGACGCTGCGCGGTGAGATG	7614
Db	1141	GCCGAGGTGTGCGCCGATCCGAGGCTTCAACCAAGGCGCGCGCGACGCTGCGCGGTGAGATG	1200
QY	7615	CAGGTGCGCCACATGCCCCGCGGCTCTCTGTGGCCGCGCCCTTCCTGGAGATTCTACGCCCCG	7674
Db	1201	CAGGTGCGCCACATGCCCCGCGGCTCTCTGTGGCCGCGCCCTTCCTGGAGATTCTACGCCCCG	1260
QY	7675	ACCGAGGACCGCGCGCTCGGTGAACAGTGAACGCGCACTGGCTCCAGAGCGGATGCGCCAGG	7734
Db	1261	ACCGAGGACCGCGCGCTCGGTGAACAGTGAACGCGCACTGGCTCCAGAGCGGATGCGCCAGG	1320
QY	7735	CTGGATACCCGAGCTGGGGTTCGCGCTTCAATCTCTGTGAACGACTTTCGCCCGGAGAGTCCCG	7794


```
Db 1321 CTGGTGAACCGAGCTGGGGGTGCGCGCTTCGATCTCGTGAACAATTGCGCCGAGAGTCCCG 1380
Qy 7795 GTGTGAGGCGCTCGGTACCGGCGCCGCACTCAAGGGCGCTGGAACCCCGACCGCTCCGGTCC 7854
Db 1381 GTGTGAGGCGCTCGGTACCGGCGCCGCACTCAAGGGCGCTGGAACCCCGACCGCTCCGGTCC 1440
Qy 7855 TGAACCTCGAGCAACCGGGATATGCTGTGACCGCCAGGTCAAGCCCGCAACAAGCTTCGGGTG 7914
Db 1441 TGAACCTCGAGCAACCGGGATATGCTGTGACCGCCAGGTCAAGCCCGCAACAAGCTTCGGGTG 1500
Qy 7915 ACCGAACAAGGCGCTGACCGCCCTTGACGAGATGACCGGGTCAACCGCGGTTCGGGAAGCC 7974
Db 1501 ACCGAACAAGGCGCTGACCGCCCTTGACGAGATGACCGGGTCAACCGCGGTTCGGGAAGCC 1560
Qy 7975 GCGGTGCTGGTGGGGGGTGGTGGCGGAGTGGCGGCAACAACGGTGGGCAACGCGGTCTGTG 8034
Db 1561 GCGGTGCTGGTGGGGGGTGGTGGCGGAGTGGCGGCAACAACGGTGGGCAACGCGGTCTGTG 1620
Qy 8035 GCCGTCAACCGAGCTTCCGAACTGGCGGCAAGATTGCCGAGACCGCGAGACCGCGAC 8094
Db 1621 GCCGTCAACCGAGCTTCCGAACTGGCGGCAAGATTGCCGAGACCGCGAGACCGCGAC 1680
Qy 8095 CGTGTGTGACGAGAGTGTGCGGACGAGTCCGCGCTTCACTTGAAACGCGCGACCGCC 8154
Db 1681 CGTGTGTGACGAGAGTGTGCGGACGAGTCCGCGCTTCACTTGAAACGCGCGACCGCC 1740
Qy 8155 GCGTTCGACCGCGCGCGGTGGCGGGGTGACGCTCCCGACCGGTGGCGAGGTGACAGTGTGTC 8214
Db 1741 GCGTTCGACCGCGCGCGGTGGCGGGGTGACGCTCCCGACCGGTGGCGAGGTGACAGTGTGTC 1800
Qy 8215 GTCCGCGCGGCGGAACCGGTATCCGAGGTCTTCAACCGATCCCGACCGGTTCGAGTGTGAC 8274
Db 1801 GTCCGCGCGGCGGAACCGGTATCCGAGGTCTTCAACCGATCCCGACCGGTTCGAGTGTGAC 1860
Qy 8275 CGTGGCGGCGGACGCGCGGAGATCTGTGCTGCTCCGCGCGCGCTGCGCCCGACCGACTGTGAC 8334
Db 1861 CGTGGCGGCGGACGCGCGGAGATCTGTGCTGCTCCGCGCGCGCTGCGCCCGACCGACTGTGAC 1920
Qy 8335 GCGCTGTGTGACCACTCGGCGGACCGCGCGCGGTGGCGGGCGCGCGCGGTGGTTCGCGCGG 8394
Db 1921 GCGCTGTGTGACCACTCGGCGGACCGCGCGCGGTGGCGGGCGCGCGCGGTGGTTCGCGCGG 1980
Qy 8395 CTGTCCCGCTTCCGGGCGCGGTATGACGAGTGTGACCGCGTTCGCGCGCGGTGTCTGACG 8454
Db 1981 CTGTCCCGCTTCCGGGCGCGGTATGACGAGTGTGACCGCGTTCGCGCGCGGTGTCTGACG 2040
Qy 8455 CGTTGCCCGGTTCGAGCTGTGAGAGAAAGATGCGCGTGTGTTTTCATGATGTGCTGT 8514
Db 2041 CGTTGCCCGGTTCGAGCTGTGAGAGAAAGATGCGCGTGTGTTTTCATGATGTGCTGT 2100
Qy 8515 CAACAGCCATCTGTTTCGGGCTGTGCTCCGCTCGCAAGCCCTTCCAGGCGGCGGACAGCA 8574
Db 2101 CAACAGCCATCTGTTTCGGGCTGTGCTCCGCTCGCAAGCCCTTCCAGGCGGCGGACAGCA 2160
Qy 8575 GGTACGGGTCGTGCGCTCGCGCGGCGCTGACCGAGCAAGTGTGCGGCTGTGACCGC 8634
Db 2161 GGTACGGGTCGTGCGCTCGCGCGGCGCTGACCGAGCAAGTGTGCGGCTGTGACCGC 2220
Qy 8635 CGTGCCTCGGTGACGAGTGTGAACTTGTGAGTGTGACGCGCCACGCGGCGCAGAGCAT 8694
Db 2221 CGTGCCTCGGTGACGAGTGTGAACTTGTGAGTGTGACGCGCCACGCGGCGCAGAGCAT 2280
Qy 8695 CGTGAAGTACATGCGGAGCCCTTCGACTGGGTGACCAAGGCCACACACATGCTCTGGA 8754
Db 2281 CGTGAAGTACATGCGGAGCCCTTCGACTGGGTGACCAAGGCCACACACATGCTCTGGA 2340
Qy 8755 CGACTCTCTGGGACATGACGACCACTTACCCCGACTTCTTGGCGCTGATGAGAGCCGGA 8814
Db 2341 CGACTCTCTGGGACATGACGACCACTTACCCCGACTTCTTGGCGCTGATGAGAGCCGGA 2400
Qy 8815 CTGCGTCAATGACGAGGATGTGAGTGTGCGCGCTCTTGGCGGTCCCGACTGATGTGTG 8874
```

```
Db 2401 CTGCTCATGACGAGGATGTGAGTTCGCGCGCTCTTGCGGTCCGACTGATGCTGTG 2460
Qy 8875 GGAAGCGGTACCTTTCGCGCGCCCGATGCGGCGCGGGTCAACGGAACCCCGCAGCGCCG 8934
Db 2461 GGAAGCGGTACCTTTCGCGCGCCCGATGCGGCGCGGGTCAACGGAACCCCGCAGCGCCG 2520
Qy 8935 GATGCTGTGGGGTTCGGAAGTGTGCGCACCGCGGCGCGGAGAGCTTCTTGGAATGTGTG 8994
Db 2521 GATGCTGTGGGGTTCGGAAGTGTGCGCACCGCGGCGCGGAGAGCTTCTTGGAATGTGTG 2580
Qy 8995 CCAACGAGAGTGGAGCAACGGGAGGATCC 9024
Db 2581 CCAACGAGAGTGGAGCAACGGGAGGATCC 2610

RESULT 3
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match 14.6%; Score 1309.4; DB 3; Length 13613;
Best Local Similarity 67.7%; Pred. No. 7.2e-206;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

Qy 2369 GGGGCGGTTCGACGCGCTTCAGACGCGCGGATGAGCGCCGAGATCCGATCCACCGTGC 2428
Db 10159 GGGGCGGATTCGGGCGGGTGGCGGTCAAGCGAGAGCGCGGCGCTCTCCAGCGGTGC 10218
Qy 2429 GCGAGTCCCGCTTGAAGTGTGAGCGCGGCGGTCAACGAGTGTGAAACCGTGCAGG 2488
Db 10219 GCGAGTCCCGCTTGAAGTGTGAGCGCGGCGGTCAACGAGTGTGAAACCGTGCAGG 10278
Qy 2489 AAGAACTGTCCCGCGCTGCGGATGATGCTGCGGCGCGGACGATGACGAAAGTGTGACG 2548
Db 10279 AAGTAATGTGCGCGGTGACGCGCGGCAACCTCGCGCGCTGACGAAAGTCCCTGACG 10338
Qy 2549 ACCGAGTGCAGGCTCCGCTCGGGGTGACCGCGCGGCGAGCTGAGCGGGTGGCCCCCGCC 2608
Db 10339 ACTCGGTGAGGAGGTGTGCGGGGTCAACGCGGCGCGGAGTGAAGGAGTGCACCGTCC 10398
Qy 2609 AGCCCGGGGAAACCGGCGCTCCCGGTACAGGTGAGTCTCCGAGCAGGTGATGTGAC 2668
Db 10399 AGGTGGGGAAGCGGCGCTCGCGGTACAGGTACAGTTCGCGAGAGATGACCTGACAC 10458
Qy 2669 GCAACTTGCAGATGAGCGGTGCGGCGCATGCTTTCGCGCGGATCCGCAACAGCTGGCG 2728
Db 10459 GCAACTTGCAGATGAGCGGTGCGGCGCATGCTTTCGCGCGGATCCGCAACAGCTGGCG 10518
Qy 2729 TCGAACCCCGACGAGGCTTCCAGCGCGTAAACCAAGTGTGTGATGCTGATGCTGCGG 2788
Db 10519 TCGAACCCCGACGAGGCTTTCAGAGCGCGTAAAGTGTGATGATGATGATGATGATG 10578
Qy 2789 CGCTCGGCGGTGTGATGACGAGCGGCGGCGCTTCCGCAACTGCGGTGCTTGC 2848
Db 10579 CGCTCGGCGGCGCTTCTCGAAGCGGTGAGGAGCTTCTGAGAGTGCAGCGCTCTCC 10638
Qy 2849 TCGAACAACTGCGGTGTCCGACCGCTGTAGTCTTCGCGAGCGGTGACGAGTGCAGC 2908
```


Db 10639 TGGGAGCTTGGCGTGTGACAGGCGCGCTGATGCTCGGAAATGTTGACGAAGTCATC 10698
 QY 2909 GGGCGGTGGGGGCTGATCTGTGAATTCGGCCATGAAGTCGACACAGTGTGATGAGCGG 2968
 Db 10699 GTCTTGCTCGCCGCGGGGTGTGAGGTTCGCGCATGAAGTCGACAGTGTGAGAGACGG 10758
 QY 2969 TTGGCCCGCCCCGCGAGATGATGTGTGAGGCGGAGCCGATCGGTGCGCCGCTTCG 3028
 Db 10759 GAGGCAACGGCCCGGAGACAGATGTAGCGAAACGAGATTGATCGGCACTGGCCTTCG 10818
 QY 3029 GCGCGATCCGCAAGAAAGTCCCGAGGTTCTTCTGAACCCGTTGAAAGGCCACGCTTG 3088
 Db 10819 GCGCGAGCTGTGAAAGCGGCGCAGGTTCTCGCGAGCGCGCGGAAAGCGGCGCTTCG 10878
 QY 3089 CCGGTGTGGCCCTGTGATCTGTGTGTTTCAAGCCCGTACAGAGGAGATCGACCGCGCG 3148
 Db 10879 CCGGTGTGTGCTGTGATCTGTGTGTTTCAAGCCCGTACAGAGGAGATCGACCGCGCG 10938
 QY 3149 AGGTCCCAACAGCGCGGCTGGCGGCGCGAGGCTGTGTTGCGTGAAGGCGAAAGCGTTGGTG 3208
 Db 10939 AGGCCCCAGAGGCGCGGCTGGCGCTCGAGGCTGCGCTCGGTGAGCGCGAAAGAGTTCTG 10998
 QY 3209 TAGACGCTCAGGCGGAGAGCCCGCGCGCGGTGCGACACCAACGCCCGGATACCGCGG 3268
 Db 10999 TAGACGCTGGGCGCGAGCGGTGTGCGTGTGCGCGCGCGCGCGCGCGCGCGCGG 11058
 QY 3269 TTGGTCAACGCTTTCAGGCGCGCGGAGAGGTAATGCGCGGTCTTCGAGCGGAC 3328
 Db 11059 TTGGTGAAGCGGCTTCAGGCGCGCGGAGAGTATCGCCGAGGCGTTTTCGCGCGGATC 11118
 QY 3329 TCGTCATGACGAGAGTGAAGATCTGTTGCGCGGCGACCAACGACTCCGCGTGTGCTTG 3388
 Db 11119 TCGTCATGACGAGAGTGAAGATCTGTTGCGCGGCGACCAACGACTCCGCGTGTGCTTG 11178
 QY 3389 GCACTGTGACCCGTCAGCAGAGTGAACAGCGGAAATGACGTCGCGCGCGGAGTGAAGC 3448
 Db 11179 GCGCGGCTCAGCAGAGTGAAGAGTGGAGCGGAAATGACGAGTTCGCGCGCGGAGTGAAG 11238
 QY 3449 CCGACGCTGTACCGGAAAGCGCGGCTCCCGCGTACCGGACGCTTCGACCGCGCGCG 3508
 Db 11239 CCGACGCTGTACCGGAAAGCGCGGCTTCGCGCGAGCGCGCGCTTCGAAAGCGCGCGCTGT 11298
 QY 3509 TTCAAGGCGCGTGAATCGGTTGCTCAAGTACTTGGCGCGGCGCGCTTCGACCGCGCTG 3568
 Db 11299 TTCAAGGCGAGCAGGAGTGTCTTTCAGTACGCGCGCGCGCGCGCGCTTCGACCGCGCTG 11358
 QY 3569 CCCAGCTGCGGAGTGCATTCGAAACAGGTTCGAGCAGCTCGCGGACCGCGCGCGCTTCAGG 3628
 Db 11359 CCGAGCTCGGAGCCTGCGCGGAAACAGGCGGAGGCGCGGAGGCGCTCCCGTTCAGCG 11418
 QY 3629 TCGAACCGACGCGGCTGTGCTTCGAAACGCGGAGTGAACGCGGAGCGCGCGCTG 3688
 Db 11419 CCGAGGCTGTGCGGCGCTTCCTTCAGCGGAGTGAAGGCGGCTGTTCGCTGACGAGCG 11478
 QY 3689 AGTTGCGGAGCGCGGACCGCGCTGTCTGTTGCGCGGAGTTCGAGATCGCGGACCGCGGAGCG 3748
 Db 11479 AGCGGAGCAGAGTGGCGGCGGCGTGTTCGCGCTTCGTCGCGCGGACGAGGCGCGCGCG 11538
 QY 3749 GTGACCTCGGCGCTGAGCGCGCGCACACAGCGCGCGCGCGCTCGCGGCTTCGCGCGCG 3808
 Db 11539 GCGAGGAGTCTGCGCGGACGCGGCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGAG 11598
 QY 3809 TCCGCGATCTCGGTGCGGTAACGCGGCTGCGCGGAGAGTGTCTTCAATCGAGCGGAC 3868
 Db 11599 CGTTTCGCGCGCGCGTGTGCGGAAAGCGCGCGGCGCGCTCATCGGAGCGCTCAATGCTG 11658
 QY 3869 CTTCGCGCGCTGTGTGCGAGAGTACAGAACGCTTTCGCTTCGCTTCGCTTCGCTTCGCG 3928
 Db 11659 GCGTGAATGTCTGCGGCGCGCGAGCGCGCGCGCGCGCGCGCTGTCGCGTGC-----GCCG 11713
 QY 3929 GTTGATCAACGACCGCGTGTGTGCGAGTCCGATGACGTCGACACCGCGCGGAGTTC 3988

Db 11714 GGTCAATTGCGCGCGCGCGGTGCGCGAGAGACGACAGATGCGGACCGCGCGGATGTC 11773
 QY 3989 CTCACTGACACCGAGAGGCGCGGTTCGAGAGGCGAGACACCTGTTCCGCGGACACTGCTG 4048
 Db 11774 GTCTGCGCGAGTGGCGGCTCGGTTCGCGAGGAGACACACCGCGCGCGGCGGCTTCGCT 11833
 QY 4049 CTGTGTCAAGCGCGAGGCTGCTCGGTGCGGAGTGTGAGTCAATCTGTTGCGAGCGGGGA 4108
 Db 11834 GTCCGCGAGCGGCGGCTGCGGCTGCGCGGAGTTCGCGGCTTCAGGCTTCGTTGAGCCCGGCGA 11893
 QY 4109 GAAATGAGGCTGTGACGACACTTCTCGCGCGCAGAGATCGCTTCGACCGCGTCAACGCTC 4168
 Db 11894 GAAATGAGGCGGAGTGTGACCGCTTCGCGCTTCAGAGACTTCATGACGAGTTCGCGGTG 11953
 QY 4169 GATGCGCGGTGCGGCTGCTTCACAGAGATGATGATGATGATGATGATGATGATGATGATG 4228
 Db 11954 GATGCGCGGTGCGGCTGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 12013
 QY 4229 GGGCGGAGCGAGATGACGAGTGAACGCGCGGTAAGTGTGCGCGAGCTCGCTGTGTAAGCGC 4288
 Db 12014 GCGGTGTGTGTGCGGACAGAGACGCGCGGAGGTTCGCGAGGTGCTTCGCGTGAAGCGGC 12073
 QY 4289 GTGTTGACCCCGGTTGTGCTCCCTGTGCTTCGCGGAAACGCTTCAGGAGAGTGAAGCCAT 4348
 Db 12074 GTGTTGACCCCGGTTTCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 12133
 QY 4349 GGGCGCGGCGACTCGCTCATCTTGTGCGTTGTGCGGATCTCGGTAACAATTGTCGG 4408
 Db 12134 GGGCGCGGCGGCTTCCTCATCTTGTGCGTTGTGCGGCGGCGGCGGCTGCGCGGCGAG 12193
 QY 4409 GCGGATGCGGAGTGTGATGAGCGCGGATCGGCTTCGCGGAGAGCGCGGCGGCGGAGC 4468
 Db 12194 GTGAAAGCGGAGTGTGAGAGGCGGAGATCGGCGGCGGAGGTTCGCGGCTGTCGCGGAC 12253
 QY 4469 CACCGCGCGCGCTTCGAAAGCGGAGTGAACCGCTTGTGTGAGGAGTGAACACTTCGCG 4528
 Db 12254 GACGCGCGCGCGCTTCGAAAGCGGAGTGAACCGCTTGTGTGAGGAGTGAACACTTCGCG 12313
 QY 4529 GTGACCGGATTCGCGCACCGGAGCTTCACCGCTGTGCAAGCCAGGAGCGTGGCGCGCTC 4588
 Db 12314 GTGCGCGAGGCTGCGCGCGCGCGCGGCTGTGACCGCGGAGCGGAGGCGGCGGCGCTC 12373
 QY 4589 GAAAGAGACTTGAACCTGTGAGTGTGCGGCGGAGATCTTTCGCGGCGCTTCACAGGATCTG 4648
 Db 12374 GAAAGTACCGCGCGGCGGCTGTGCGGAGCTTCGCGGAGTGTGCGGCGGCGGAGG 12433
 QY 4649 TCGGCGCGCACAGGTGACCGGACGATTCGCGCGGCTGTGCGGAGTGAACGCGCGGAGC 4708
 Db 12434 GCGGCGCGCACAGGTGAGCGCGGACGAGCGGAGTGTGCGGAGTGAACGCGCGGCGGAGC 12493
 QY 4709 GTGCTCGGAGTGAACGAGCCGCTGCGCGGAGTGTGCAAGTGTGCAAGACCGGAGTGAAGTTC 4768
 Db 12494 CTGCTCGGAGTGAAGTGTGCGGAGTGTGCGGAGTGTGCAAGTGTGCAAGACCGGAGTGAAGTTC 12553
 QY 4769 GAGCGAGCTGCGCGGAGTGTGCGGAGTGTGCGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 4828
 Db 12554 GATTCAGCGGAGTGTGCGGAGTGTGCGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 12613
 QY 4829 GGTGACGTGACCGGCGCGGAGACAGTTCGAGGAGGAGCGGAGTGTGCAAGTGTGCAAGTTC 4888
 Db 12614 GGTGAGGCGCGGAGTGTGCGGAGGAGTGTGCGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 12673
 QY 4889 GCAATGCGGATCCCGGACGAGTTCGCGGAGTGTGCGGAGTGTGCAAGTGTGCAAGTTC 4948
 Db 12674 GCAATGCGGATCCCGGACGAGTTCGCGGAGTGTGCGGAGTGTGCAAGTGTGCAAGTTC 12733
 QY 4949 GCGGTTGAGGAGCGGAGTGTGTTTCAAGGAGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 5008
 Db 12734 GCGGTTGAGGAGCGGAGTGTGTTTCAAGGAGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 12793
 QY 5009 GTGCGGAGTGTGCGGCGCGGACGAGTGTGAGGAGGAGTGTGCAAGTGTGCAAGTGTGCAAGTTC 5068
 Db 12794 GTGATGCGGTTGGGCGCGCGCACGAGAGGCGGCTGTGTGAAGGCGGCGGCGCGGAG 12853

FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/657,440
CURRENT FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: CIP OF 09/141,908
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 1565
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-657-440-24

Query Match 8.28; Score 729.4; DB 3; Length 1565;
Best Local Similarity 68.18; Pred. No. 7.5e-111;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

2369 GGGGCGGTCGACGCGCTTCAGACGCGCGGATTCAGCGCCCGATGCTGATCCACCGCTCG 2428
1530 GGGGGGATCGGGGCGGGTGCAGGTGCGAGGAGCGCGGCTCTCCACGCGCTCC 1471
2429 GCGACGTCGCGCTTCAGATGCTGAGCGCGGCGGTCAACGACTGCTGAAAACCGTTCAGG 2488
1470 GCGGCGTCGCGCTTCAGACTGCTTCAGCGCGCGGTGACGACTGATCGAAGCGGTCCAG 1411
2489 AAGACTGTCGCGCGCTGCGGATTCAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2548
1410 AAGACTGTCGCGCGCTGCGGATTCAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1351
2549 ACCGATGTCAGGCTTCGCTTCGCGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2608
1350 ACCGATGTCAGGCTTCGCTTCGCGGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1291
2609 AGCGCGGAGAAACCGGCTTCGCGGCTGACGAGTTCGCGGCGCGCGCGCGCGCGCGCGCGCG 2668
1290 AGGTCGCGGAGAAACCGGCTTCGCGGCTGACGAGTTCGCGGCGCGCGCGCGCGCGCGCGCG 1231
2669 GCGACCTGCGGATGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2728
1230 GCGACCTGCGGATGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
2729 TCGACACCG 2788
1170 TCGGCG 1111
2789 GCGTCG 2848
1110 GCGTCG 1051
2849 TCGACACCG 2908
1050 TCGGCG 991
2909 GCGCGCTGCGGCTGCGGCTTCGCGGCTGCGGAGTTCGACCGAGGTCGATGAGGCGG 2968
990 GTCCTGCG 931
2969 TTGCG 3028
930 GAGGCGACCG 871
3029 GCGCGCATTCGCGAGGAGTTCGCGGAGTTCCTTCGACCGCGCTTCGAAAGCGCGCGCTTG 3088
870 GCGCGCAGCTGCTGGAAGCG 811
3089 CCGGTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3148
810 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
3149 AGGTCCACAGGCG 3208

750 AGGCGCGAGAGCGCGGCTGCGGCTTCAGAGGTGCGCTCGGTGAGCCCGAAGATTCGTG 691
3209 TAGACGATGAGGCGGAGCG 3368
690 TAGACGATGAGGCGGAGCG 631
3269 TTGATGACGCTTCGAGCG 3328
630 TTGATGACGCTTCGAGCG 571
3329 TCGTCCATGACCGAGGTCGAGGATCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3388
570 TCGTCCATGACCGAGGTCGAGGATCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
3389 GCACTGTCACCGCGTACGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 3448
510 GCGCGCGTTCACGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAG 451
3449 CCGACGCTGTCAGGAGAGGCG 3508
450 CCGACGCTGTCAGGAGAGGCG 391
3509 TTGACGCGCTGATCGTGTGCTTCAGTACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3568
390 TCGAGCGGAGCGAGGCTGCTTCAGTACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
3569 CCGACGCTGTCAGGAGAGGCG 3628
330 CCGAGCTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 271
3629 TCGAACCGAGCGCGATCTGCTTCGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 3688
270 CCGAGCTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 211
3689 AGTTCGCGAGCGCGAGCG 3748
210 AGCGGAGCGAGGTCGCGGCG 151
3749 GTGACCTGCGGCTGAGAGCG 3808
150 GCGAGGTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 91
3809 TCCGCGATTCGCGGTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 3859
90 GCTTCG 40

RESULT 7
US-09-793-708-24/c
Sequence 24, Application US/09793708
Patent No. 6902913
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT MARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002121
CURRENT APPLICATION NUMBER: US/09/793,708
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: US 09/657,440
PRIOR FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: US 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/846,247
PRIOR FILING DATE: 1997-04-30
PRIOR APPLICATION NUMBER: US 60/134,990
PRIOR FILING DATE: 1999-05-20

/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 24
/ LENGTH: 1565
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-793-708-24

Query Match 8.2%; Score 729.4; DB 3; Length 1565;
Best Local Similarity 68.1%; Pred. No. 7.5e-111;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGCGGATCGACGCGCTTCAGACGCGCGATCAGGCGCCGATGCTGATCCACCGCTCG 2428
DB 1530 GGGGGGATCGGGGGGGGGTGGGGGTGCGAGCGAGAGAGCGGGGCTCTCCACGGCTCC 1471
QY 2429 GCGACGTCCCGTTGAGTTGTTGAGCGCGGCGGTCAACGATGTTGAAACCGTGCAGG 2488
DB 1470 GCGGCGTCCGCTCGACGTGGTTCAGGCGGGCGGTGACGACTGATGAAAGCGTCCATG 1411
QY 2489 AAGAACTCGTCCCGCGCTCGCGATGATGCTGCGCGCCGACGCTGACGAAAGTGTGAGCG 2548
DB 1410 AAGTACTCGTCCGCTCGTGCAGCGCGCCACCTGCGCGCGCTGACGAAAGTCCGTCAGC 1351
QY 2549 ACCGATGCGAGCTCCGCTCGCGGTGACCCGCGCGCGACGTAAGCGGTGCGCGCGCC 2608
DB 1350 ACCGATGCGAGCGGTGCTCGGGGTTCACGCGCGCGCGCATGTAAGCGGTGCGCGCGCTC 1291
QY 2609 AGCGCGGAGAAACCGGCTTCGCGGTACAGGTAGACGTTCGACGAGTGTGATCTGACG 2668
DB 1290 AGGTTCGCGGAGACCGGCTTCGCGGTACAGGTACAGTTCGCGCGAGATGACCTGACG 1231
QY 2669 GCGACCTGCGGAGATGCGCGGTGCGCGGTGCTGTTCCGCGCGGATTCGCGACAGCTGCG 2728
DB 1230 GCGACCTGCGGAGATGCGCGGTGCGCGGTGCTGTTCCGCGCGGATTCGCGACAGCTGCG 1171
QY 2729 TCGACACCGCGGACGAGCTTCGCGGTACAGGTAGACGTTCGCGAGTGTGATGCGCGG 2788
DB 1170 TCGGCGCGCGGTGCGCGGTGCGCGGTGCTGTTCCGCGCGGATTCGCGAGTGTGATGCG 1111
QY 2789 CGCTCGCGCGGTGCTGAGTGCAGAAACCGGCGGAGCGCTTCGCGACACTGCTGCTGCG 2848
DB 1110 CGCTCGCGGACCGGCTTCGCGAGGCGGTGAGGCGCTTCGCGAGCTCGCGCGCTCTCG 1051
QY 2849 TCGGACACCTGCGCGGTGCTTCGCGGTGCTGCTGCGGACGCGGTGACGAAAGTGCAGC 2908
DB 1050 TCGGCGGAGTTCGCGGTGCTGCGCGGTGCTGCTGCGGATGTTGACGAAAGTGCAGC 991
QY 2909 GCGCGGTGCGGCTGAGTCTGTTGAGTTCGCGGTGCGGATGCGACGAGTGCAGTGCAGC 2968
DB 990 GTCCTGCGCTGCGCGGTGCTGTTGAGTTCGCGGATGCGACGAGTGCAGGTCGAGC 931
QY 2969 TTGCGCGCGCGCGCGGACGATGATGTTGAGGCGCGGACCGGTCGCGCGCTTCG 3028
DB 930 GAGGCGACCGCGCGGAGCGAGTGTGAGGCGGAGGTTGATTCGCGGACCTCGCGCTG 871
QY 3029 GCGCGGTGCGGACGAGATTCGCGGAGTTCCTGTAACCGCTTCGAGGCGCGGCTTCG 3088
DB 870 GCGCGGAGTTCGAGGAGGCGCGGAGTTCGCGGACCGGCGGAGGCGCGCTTCG 811
QY 3089 CCGGTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3148
DB 810 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 751
QY 3149 AGGTTCGCGACGCGCGGTGCGCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3208
DB 750 AGGCGCGGAGGCGCGGCTGCGCGCTGCGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTG 691
QY 3209 TGAAGCTGCGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268
DB 690 TGAAGCTGCGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
QY 3269 TTGCTCAAGCTTCAGCGCGCGCGGAGGATCATCGCGGTGCGGCTTCGCGCGCGC 3328

DB 630 TTGCTGAGGCGCTCCAGCGCGCGGAGGATCATGCGCGAGGCTTCGCGCGGATTC 571
QY 3329 TCGTGCATGACCGGAGTGAAGATCTGCTGCGCGGACCAACGATTCGCGGTGCTG 3388
DB 570 TCGTGCATGACCGGAGTGAAGATCTGCTGCGCGGACCAACGATTCGCGGTGCTG 511
QY 3389 GCACTGCTGACCGGATTCGCGGAGTGAAGATTCGCTGCGCGGACCAACGATTCGCG 3448
DB 510 GCGCGGTGACCGGAGTGAAGATTCGCTGCGCGGACCAACGATTCGCGGTGCTG 451
QY 3449 CCGACGCTGACCGGAGTGAAGATTCGCTGCGCGGACCAACGATTCGCGGTGCTG 3508
DB 450 CCGACGCTGACCGGAGTGAAGATTCGCTGCGCGGACCAACGATTCGCGGTGCTG 391
QY 3509 TTCAAGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3568
DB 390 TCGAGCGGAGGAGGAGTTCCTTCCAGTACGCGCGCGCGCGCGCGCGCGCGCG 331
QY 3569 CCGACGCTGCGGAGTGAAGATTCGAGTGAAGATTCGCTGCGCGGACCAACGATTC 3628
DB 330 CCGAGCTTCGCGGAGTGAAGATTCGAGTGAAGATTCGCTGCGCGGACCAACGATTC 271
QY 3629 TCGAACCGAGCGCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3688
DB 270 CCGAGGTCGTCGCGCGGCTTCCTGCGCGGAGTGAAGATTCGCTGCTGCTGCTG 211
QY 3689 AGTTTCGCGAGCGCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3748
DB 210 AGCGGAGCGAGGAGTTCGCGCGGCTTCCTGCGCGGAGTGAAGATTCGCTGCTGCTG 151
QY 3749 GTGACCTGCGCGGCTGAGCGCGCGGACCGCGCGCGGCTGCGCGGTGCTGCTG 3808
DB 150 GCGAGGCTTCGCGGAGTGAAGATTCGCGCGGCTTCCTGCGCGGAGTGAAGATTC 91
QY 3809 TCGGAGATTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3859
DB 90 CCGTTCGCGCGGAGTTCGCGGAGTGAAGATTCGCGCGGCTTCCTGCGCGGAGTTC 40

RESULT 8
US-09-105-537-9/C
/ Sequence 9, Application US/09105537A
/ Patent No. 6265202
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and plikromycin
/ FILE REFERENCE: 600,438US1
/ CURRENT APPLICATION NUMBER: US/09/105,537A
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 9
/ LENGTH: 1458
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-105-537-9

Query Match 8.1%; Score 722; DB 3; Length 1458;
Best Local Similarity 69.5%; Pred. No. 1.2e-109;
Matches 980; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 2395 GCGATTCAGGCGCGGATGCTGATCCACCGCGCGGACGTCGCTGCTGCTGCTGCTG 2454
DB 1455 GCGGAGAGGCGCGGCTTCCTCCAGCGCTTCGCGGAGTGCCTCCAGCTGCTGCTG 1396
QY 2455 CCGGCGGTGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2514
DB 1395 GCGGCGGTGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1336

QY	2515	GATGCTGCGAGCCGCACTGTCAGAAAGTGTGAGCAGCCAGATGCAGGCTCCGGTTCGAGGT	2574	
Db	1335	CACCTGCGCGCGCGGCTCGACGAAGATCCCTTGA	CGACTTCGTGTGGAGAGGTGTTCGAGGAGT	1276
QY	2575	GACCCGCGCGCGGACGATGAGCGGGGTGGCCCCCGCAGCGCCGCGGGGAAAACCGGCTCCCGGAT	2634	
Db	1275	CACGCGCGCGCGGATGATGCGGGGTTCGGCGCT	CCAGGTCGGGGAAAGCCGGCTTCGCGAT	1216
QY	2635	CAGGTAGACGTCTCCGAGCAGAGTCGATCTGACCGCCACCTCGGATGGCGAGTCGGGCG	2694	
Db	1215	CAGGTAGACGTCGCGCAGAGATGCAGCTTGACCGCGACCTGCGGGGTGGCGCGGTGGGCGCG	1156	
QY	2695	CATCGTTTCCGCGCGGATCCGCAACAGCTGGGCGCTCGACACCCCGACGCGAGGCTCTCCAA	2754	
Db	1155	CATGCTGCGCGGACTTGTATCCGACAGATTCGCGCTCGCGCCCGGTGTGCGCAGGCTGTTCAG	1096	
QY	2755	CGCGGTACCCAGGTCGCGTGTGATGACCGCGGGGTCCGCTCGCGCGCGGATGATGACGAACCG	2814	
Db	1095	GCGGTAGCGGTAGTGTGATGTGAGTCCGCGGGGTGGCTCGCGGACCCGCTCTCGAAGGC	1036	
QY	2815	GCGCAGAGCGCTCCCGCAACTCGGTCGTTCCGCTCCGACCAACTGCGCTGTCCCGACCC	2874	
Db	1035	GTTGAGGGCGCTCTGTGAGCTCGGCCCGCTCCCTCGTGGGACAGCTTGCCTGTGACGAGCC	976	
QY	2875	GCTGTAGTCTCTCGCGGACGGTGACGAGTGCAGAGCGCGCGGTGGCGGCTCGAAGCTGTTGAG	2934	
Db	975	GCTGTAGTCTCTCGCGCAAGTTGACCAAGATTCATGTCCTGTGCTCGCCCGCGGTGTGTGAG	916	
QY	2935	TTCCGCGGATGATGATGACCAAGTTCGATAGGCGGTTGGCGCGCCCGCGCGAGATGATGTG	2994	
Db	915	GTCGCGGATGATGATGACCAAGTTCGAGCGAGCGGGGAGGACCGGCCCGGAGACAGATGTA	856	
QY	2995	GTTTGAAGCGCAGCCGAGCCGCGGTGCGCGCCGTTTCGCGCGCATTCGCGAGAGTCCCGAG	3054	
Db	855	GCGGAGCGCGAGGTGTGATTCGCGAATCCGCGCTCGCGCGCGCAGCTGTGAGAGCGGCGCAG	796	
QY	3055	GTTCTTCTGTACCGGTTTCGAAGAGCGCCACGCTTGCCTGGTGTGCTGTATCTCGTGTT	3114	
Db	795	GTTCTTCGCGAGCGCGGCGGAAGGCGGCGCTTCTGTGCGGTGTGTGTGTGTCTCTCGTGC	736	
QY	3115	GTTTCAGCCCGTACAGAGAGGTACCGACCGCGCGAGAGTCCACAGCGCGGGCTGGCGCG	3174	
Db	735	GTTTGAAGCGGTGAGCGAGAGGTGCGGATGGCGTGGCGAGCGCCCGAGAGCGGGCTGGCGCTC	676	
QY	3175	CAGCGCTGTGTGAGTGAAGGCGGAGAGGCGTGTGTGTGACAGGTCAAGAGGCGAAGCCCGCG	3234	
Db	675	CAGGAGTGCCTGCTGTGAGCGCGAAGAGATTCGTGTGACGATGCGGCGCGCAGCGCTGTGCTC	616	
QY	3235	GCGCGCGTGCAGACCAACAGCGCCCGGATACCGCGGTTTGGTCAACGTTTCAGGCGCGCGGA	3294	
Db	615	GGTGGGTTGCGCGCGCAGGCTCCCGAGGCGCGGGGTTTGGTGTGAGCGGCTCAAGCGCGCGGA	556	
QY	3295	GAGGTACATCGCGCTGTGGGTTTCTCCGACGGGACCTTCGTGTGATGACCGAGGTGAGGATTC	3354	
Db	555	GAAGTACATTCGCGCGAGGGGTTTCCCGCGGTTATCTTCGTGATGACCGAGCATGTGC	496	
QY	3355	GTTTGCGGGACCAACAGCTCGCGCGTGTGATTTGGCACCTGTGCACCCGTCACGAGAAATG	3414	
Db	495	GTTTGCGGGCTGTGAGGCGCGAGCGGTCGTGAGCGGGCGCGGTGTGCACCGAGCGAGAGTG	436	
QY	3415	ACAGCGGAACATGCAAGTCGCGGCGCGGGGTGAGCCGACGCTGTACCGGAAAGCGGGCTC	3474	
Db	435	GCGCGCGAAATGTCAGATTCGGGCGCGGGGTGAGGCGCGACGCTGTACCGGAAAGCGGGCTT	376	
QY	3475	CGCGCGGTACCGCAGCTCCAGCGCGCGCGCGCGGTTCAGCGGCGCTGATCTGTGTCTCCA	3534	
Db	375	CTGTGCGAGCGCGCGCTCGAAGACGCGCGCGCTGTTCGACCGGAGACAGAGGTGTTCCTCA	316	
QY	3535	GTACTTGGCCCGCGGGCGCTGTGCACCGCGCGGTCCAGCTTGGCGGGATTCGATCGAACG	3594	
Db	315	GTAACGCGCGCGCGGGCGCGGTCTCGACCGCGGGTGCAGACCTCCGGGACCTTGCCTCCAGAG	256	
QY	3595	GTCGACGAGCTCGCGCAACCGCGCGCGGTTCAGGTGACCGACCGCGCATCTGTCTCAA	3654	

Db	255	GGCGAGAGAGCGCCCGGAAAGCGTCCCGGTCCAGCCGCCAAGTCTGTGGCCGGGCTTCCTCCAG	196
Qy	3655	CGGGGTGAACCGGCGAGGCGCCGTGATGAGGCGGTGATTTCGGCGAACCGGACCGGCTGTCTG	371
Db	195	CGGGGTGAAGGGGCGTGTTCGCGTAGCGACCGGCGAGCGGAGAGATGAGCGGGCGAGTCGT	136
Qy	3715	TTCCGCGGATGTCCGATGCGGACCGGTCAGGCGGAGTGAACCTTCGCGCTGAGCGGCCCGCAC	377
Db	135	TCCGGCTCTGTCCGGCGGACGAGGCGCGCCGGCGGCGAGGGTCTGAGCGACCGGCTGGAC	76
Qy	3775	CACGGCGGAGCGGTTCGGCGGTGGGTCGTGTC	3804
Db	75	CGCGGCCCCAGATCGGCTCGGGGTGGC	46

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	Best Local Similarity	Score	DB 2	Length
Matches 1118; Conservative	7.8%;	66.0%;	701.4;		3756;
			Pred. No. 3.2e-106;		
			Matches 526; Indels	51;	Gaps
				57;	
QY	7330	CGGAGCGCTTCATGCGCATGGAGTGGCGCGCTGGGCGCGGAGTTGCAGATGACC			3898
Db	42	CGGAGGAAATTATATACCAACGCGCATGGCGCGCGCTGGGCGGCACTCGAATGATC			101
QY	7390	CGGCGTCTTACTGGGGGCTTGGGTGCCAAGCGCATCTGTACTGATCTCTGTCCGGA			7449

```

Db      102 CGGCGCTGCACTGGGGTTACGGCAGCAAGCCGCACTTACCCGATGCTGTGGCA 161
Qy      7450 CGGAGCAGACGACCCCTGACCTGTGTACGAAAGTTTGGAGGGCGCGCGAGACGGGAG 7509
Db      162 CAGACGACGACGACCGGAGCGCGCTGCTGATGCGAGTCCGGTGTGCGGCG----- 216
Qy      7510 GCCAGTGGGCGGAGAGTGGGTGTGTGTGACCAACCGGACCGCGCGGAGTGTGCGC 7569
Db      217 -----GAGAGACCGGAGAGTGGGTGTGTGTGACCAACCGCGCGGAGTGTGAC 272
Qy      7570 GATTCGGGGCTTCAACCCAGCGCGCGCGGAGCGCTGCGCGGTGATGAGTGGCCACTGC 7629
Db      273 GACCCCGCGTTCACCCCGCGCACCGGACGCAACCGGATGATGCGGCGCGCGCGCG 332
Qy      7630 CGGCGCGGCTCTCGAGGCGGCGCGCTTCCGAGGAGTTTCAACCGCGGCAACGAGAGCGCGCG 7689
Db      333 CACACCGCGGAGTGGGCGCGAGCGGTTCGAGGAGCTTGCACCGCGGTCTTGGAGAG----- 387
Qy      7690 TCGGTGACAGTGGACCGCGGACTGGCTTCAGCAGCGGTGCGCGAGGCTGTGACCGAGCTG 7749
Db      388 -GGAGGGTCCCGACGTCGGGAGAACTGGCGGAGAGCTTCCCGGTCTGCTCCCGGCGCG 446
Qy      7750 GGGTCGGCTTGCATCTCTGTGAAAGACTTTCGCGGAGAGTCCCGGTGTGGCGCTGCGT 7809
Db      447 GCGCGCGGCTGGACCTGTGTGCGGAGCTTGTGCGGAGGTAACGGGTGACAGGCGATGACC 506
Qy      7810 ACCGCGCGGCACTCAAGGGGAGGAGACCCGACCGTTCCTGCGGTCTGGAACCTCGGCGACC 7869
Db      507 GCGGTGCTGGCGGCGAGCGGAGTGTG-----CTGCGCGGCGCGCGGTGAGAGCGCC 554
Qy      7870 CGGATATGCTTGGACCGCGGAGTCAAGCGCGGCAACAGCTCGCGGTGACCGAGCGGCTG 7929
Db      555 CGGTGAGCTTGGACCGCGGAGTCAAGCGCGGCAACAGCTCGCGGTGACCGAGCGGCTG 614
Qy      7930 ACCGCGCTCGACGAGATGACGCGGTGACCGCGGTGCGGAGCGCGCGGTGCTGTGGGG 7989
Db      615 GCGGCACTGCGCGCGGACCGCGGACTGCGGCGCC-----TGTTC 653
Qy      7990 GTGGTGGCGGAGCTGGGCGGCAACAGGTGGGCAACGCGCTCTGTGCTGTACCGAGCTT 8049
Db      654 GCGGCGGCGCGAGATGACCGGAGCAACGCTGTGACGCGGTCTGTGCGGTCTGTGCGCGGAA 713
Qy      8050 CCGGAACTGGCGGAGCACTTTCGCGAGCAACCGGAGACCGCGAGACCGCGTGTGACGAG 8109
Db      714 CCGGGGCTGGCGGAGTGGAGTGGCGGAGACCGCGCGCGCGCGGAGACCGGTGCGCGAG 773
Qy      8110 GTGTGCGGAGCAAGTCCCGGCGTTCCTTGAACCGCGGACCGCGCGGTGAGACCGCGCG 8169
Db      774 GTGTGCGGCTGACCGCGGACTTGGACCTTGAACCGGCGGACCGCGGAGGTCGCG 833
Qy      8170 GTGGGCGGGGTGCACTCCGAGCGGTGGCGGAGTGAAGTGTGCTGTGCGCGGCGGAGC 8229
Db      834 CTGCGGAGAGCACTGTATCGCGGAGGAGGAGAGTGTGTCTGTGCTGTGCGCGGCGGCAAC 893
Qy      8230 CGTATCCGAGGCTTTCACCGATCCGAGCGGTTGAGAGTGTGAGCGTGTGCGCGGAGCGCC 8289
Db      894 CGGACCCGAGGATCTTTCGCGAGCGGAGCGCTGTGAGTGTGAGCGCGCGGAGCGCGGAC 953
Qy      8290 GAGATCTGTGTCTCCGCGCGGCTGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGC 8349
Db      954 CGGCGCTGTGTGCGGACATGCGCGGCAACCGCGGCA-----GGTGAAGAGAGTGTGACCGG 1010
Qy      8350 CTGGGCAAGCGGCGGCTGGGAGCGCGCGCGCGGTGTGCGCGGCTGTGCGGCTGTGCGG 8409
Db      1011 CTGCGGACCGGCGGAGTGGGCGCGCGCGGAGCGGCGGCGGCTGTGCGGAGTGTGCGGAG 1070
Qy      8410 CCGGTATGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8469
Db      1071 CCGGTGTGTGCGGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1130
Qy      8470 CTGTAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8529
Db      1131 CTGTAGAGATTCGCGGATGCGGCTGTGTCTCTCTCATGTGCGGAGAGAGGCGGCTT 1190

```

```

Qy      8530 CGGCTGTGTGCGGCTGCGAAGCGGCTTTCAGAGGCGCGCGGACAGAGGTACGGGTGCGC 8589
Db      1191 CGGCTGTGTGCGGCTGCGAAGCGGCTTTCAGAGGCGCGGCGGAGGAGTGTGCGGTGTGCGC 1250
Qy      8590 CTGCGGCGGCTTTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8649
Db      1251 GTTCCCGGCGGCTTTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1310
Qy      8650 CGAGCTGTGAACTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8709
Db      1311 CGAGCTGTGAACTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1370
Qy      8710 GACCTGTGAACTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8769
Db      1371 GACCTGTGAACTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1430
Qy      8770 GCAGACCACTTTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8829
Db      1431 GCAGACCACTTTCAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1490
Qy      8830 GATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8889
Db      1491 CATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1550
Qy      8890 CGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 8949
Db      1551 CGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1610
Qy      8950 GAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 9009
Db      1611 GAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1670
Qy      9010 GCACCGGAGAGATCC 9024
Db      1671 GCACCGGAGAGGCGCC 1685

```

```

RESULT 10
US-09-105-537-7/c
; Sequence 7, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FaSTSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-7

Query Match      6.7%; Score 595.4; DB 3; Length 1248;
Best Local Similarity 69.2%; Pred. No. 6,9e-89;
Matches 812; Conservative 0; Mismatches 361; Indels 0; Gaps 0;
Qy      3926 CGCGTTGATCAAGCAGCGCTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3985
Db      1173 CGCGGTCAAGTTGCGGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1114
Qy      3986 GTTCTTCACTGAGACAGGAGGCGCGGTGTGAGGAGGCGAGACACCGTTCGCGGAGCACTC 4045
Db      1113 GTTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1054
Qy      4046 GGTGTGTGCAAGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4105

```


Db 1053 GGTGTGCGAGCGAGCGGCGTGGCGCTCCGCGGTACCGCTCCAGCTGTGGACGCCGG 994
Qy 4106 GAGAGATAGAGCGCTGTGCGACGACCTTCTCCGCGCGAGAGATGCCCTTGACCGCGTACG 4165
Db 993 CGAGAACTAGCGCGCGGTGTGACGCGCTTGGCGCTTTAGACCTCCATGACGAGGTGCG 934
Qy 4166 GTCGATCCCGGTGGCGGTGCGGTCCGTCACACAGATGATCACTGTAGTGTCTCTCTC 4225
Db 933 GTGATATCCCGGTGTGCGCTGTGCTGATCTGCAACATCACTGATCTGTGTGTGAGGCC 874
Qy 4226 GTGCGGCGGAGCGAGTGCACGCTGACCGCGGTACCTGCGCGACCTCGTGTGAG 4285
Db 873 GTGCGCGGTGTGCTGCGCGACGAGAGCGCGGAGAGGTCCGCGAGGTCTCGCGGTGAGC 814
Qy 4286 CGCGTGTGACCGCGGTGTGTCTCCCTGTGTCTCGCGAGACCGCTTCGAGGAGGTGAGCCC 4345
Db 813 GCGCTGTGTGCGCGGTTCGCGTGTGATGACTCGGAGAAACGCGTGCAGGAGGTGAGGACC 754
Qy 4346 CATGCGCGCGCGCGACTCGCTCATCTTGTGCGGTGTGCGGTGCGGTGACCACTTGTG 4405
Db 753 CATGCGCGCGCGCGCGCTGCTCATCTTGTGCGGTGTGCGCGCGCGCGCGCGCGCGCG 694
Qy 4406 CGGCGCGATGCGAGATGTGTCATGCGCGGTGCGGTGCGCGCGCGCGCGCGCGCGT 4465
Db 693 CAGGTGCGAGCGAGATGTGTCATGCGCGGTGCGGTGCGGTGCGCGCGCGCGCGCGT 634
Qy 4466 GACCAACCGCGCGCGCGCTGCGAGCGGTGACCGCGGTGCGGTGCGGTGCGGTGCGGT 4525
Db 633 GACCAACCGCGCGCGCGCTGCGAGCGGTGACCGCGGTGCGGTGCGGTGCGGTGCGGT 574
Qy 4526 GCGGTGCG 4585
Db 573 GCGGTGCG 514
Qy 4586 GTGCGAGAGAGCTGACCTGTGTGTGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4645
Db 513 GTGCGAGAGAGCTGACCTGTGTGTGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 454
Qy 4646 TGTGTGCG 4705
Db 453 GCGGTGCG 394
Qy 4706 GACGTGTGCG 4765
Db 393 CACTGTGTGCG 334
Qy 4766 TCCGAGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4825
Db 333 GCGGTGCG 274
Qy 4826 ACCGTGTGCG 4885
Db 273 GCGGTGTGCG 214
Qy 4886 GATGAGTGTGCG 4945
Db 213 CAGCGCATGCG 154
Qy 4946 CCGCGGTGTGCG 5005
Db 153 GCGCGGTGTGCG 94
Qy 5006 CCGGTGTGCG 5065
Db 93 GCGGTGTGCG 34
Qy 5066 GAGATGCG 5098
Db 33 GAGATGCG 1

Sequence 21, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
EARLIER FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 5970
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-21

Query Match 5.8%; Score 519.2; DB 3; Length 5970;
Best Local Similarity 70.1%; Pred. No. 2.5e-76;
Matches 698; Conservative 0; Mismatches 298; Indels 0; Gaps 0;
Qy 4104 GGGAGAGATGAGGCGCTGTGCGACGACCTTCTCCGCGCGAGATGCGCTGACCGCGTCA 4163
Db 1 GGGAGAGATGAGGCGCGGCTGTGCGACGACCTTCCGCGCTTACAGAACCTTCATGACGAGGTG 60
Qy 4164 CGGTGATGCGCGGTGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4223
Db 61 CGGTGATGCGCGGTGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 120
Qy 4224 TCGTGTGCGGAGAGAGTGTGCGAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4283
Db 121 CGGTGTGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 180
Qy 4284 AGCGGTGTGAGCGCGGTGTGCTCCGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4343
Db 181 SCGGGTGTGCGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 240
Qy 4344 CCGATGCG 4403
Db 241 CCGATGCG 300
Qy 4404 TCCGCGCGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4463
Db 301 GCGAGGTGTGAGCGCGGTGTGAGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 360
Qy 4464 GTGACACCG 4523
Db 361 GTGACAGAGCG 420
Qy 4524 TCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4583
Db 421 TCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 480
Qy 4584 GCGTGTGAGAGAGCTTACCTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4643
Db 481 GCGTGTGAGAGAGCTTACCTGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 540
Qy 4644 GCTGTGCG 4703

Db 541 CAGGGGCGGCGCCAGAGGTGAGCGCCAGACGCGCCAGAGGTGCGGGGTGTGACCGCGCGG 600
Qy 4704 GCGACGTGTGTCGGGTGTGACCAAGACCGGTGCGCGGGTGTGACGTGTGAGAAACACCGGTGTG 4763
Db 601 GCGACCTGTGTGCGGTGTGAGGTGTGCGGTGTGCGGTGTGCGGTGTGCGGTGTGCGGTGTG 660
Qy 4764 AGTCGAGCAGGTGCGCGGTGTGCGCGGTGTGCGCGGTGTGCGCGGTGTGCGCGGTGTGCGCGGT 4823
Db 661 AGGCGCATTCAGCGCGCATGT 720
Qy 4824 TCACCGGTGACGTGACCG 4883
Db 721 TCACCGGTGAGCG 780
Qy 4884 GCGATGACGTGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4943
Db 781 GCGACGCGCATGTGCG 840
Qy 4944 GCGCGCGCGGT 5003
Db 841 GCGCGCGCGGT 900
Qy 5004 TCGCGGTGCGCGGT 5063
Db 901 GCGCGGTGCGCGGT 960
Qy 5064 CCGAAGATGCGCGCATGT 5099
Db 961 CCGAAGATGCGCGCATGT 996

RESULT 12

US-09-141-908-11
; Sequence 11, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-11

Query Match 5.8%; Score 519.2; DB 3; Length 5970;
Best Local Similarity 70.1%; Pred. No. 2.5e-76;
Matches 698; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

Qy 4104 GGGGAAAGTGGCGGT 4163
Db 1 GGGGAAAGTGGCGGT 60
Qy 4164 CGGTGATGCGGT 4223

Db 61 CGGTGATGCGGT 120
Qy 4224 TCGTGTGCGCGGAGCGAGT 4283
Db 121 CCGTGTGCGGT 180
Qy 4284 AGCGGT 4343
Db 181 SCGGGT 240
Qy 4344 CCATGTGCGCGGCGCATGT 4403
Db 241 CCATGTGCGCGGCGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
Qy 4404 TCGGCGCGATGT 4463
Db 301 GCGAGGT 360
Qy 4464 GTGACACG 4523
Db 361 GTGACACG 420
Qy 4524 TCGGCGGTGACCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4583
Db 421 TCGGCGGTGCGCGAGGT 480
Qy 4584 GCGTGTGAGAGAGCTTGT 4643
Db 481 GCGTGTGAGAGAGCTTGT 540
Qy 4644 GCGTGTGCG 4703
Db 541 CAGGGGCGGCG 600
Qy 4704 GCGACGT 4763
Db 601 GCGACCTGT 660
Qy 4764 AGTCGAGCAGGTGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4823
Db 661 AGGCGCATTCAGCGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Qy 4824 TCACCGGTGACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4883
Db 721 TCACCGGTGAGCG 780
Qy 4884 GCGATGACGTGCGCGGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4943
Db 781 GCGACGCGCATGTGCG 840
Qy 4944 GCGCGCGCGGT 5003
Db 841 GCGCGCGCGGT 900
Qy 5004 TCGCGGTGCGCGGT 5063
Db 901 GCGCGGTGCGCGGT 960
Qy 5064 CCGAAGATGCGCGCATGT 5099
Db 961 CCGAAGATGCGCGCATGT 996

RESULT 13

US-09-657-440-21
; Sequence 21, Application US/09657440
; Patent No. 6509455

; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li

QY 4344 CCCATGCGCGCGGCACTGCTCATCTTGCCGTTGGTGCAGATCTGCTGACCACTTG 4403
 Db 241 CCATGCGCGCGGCGGCTGCTCATCTTGCCGTTGGTGCAGATCTGCTGACCACTTG 300
 QY 4404 TCCGCGCGGATGCGGAACTTGTGATGAGCCCGATCCGTTCCGCTGACGAGCCCTGCTG 4463
 Db 301 GCGAGGTGCAAGCGGAACTTGTGATGAGCCCGATCCGTTCCGCTGACGAGCCCTGCTG 360
 QY 4464 GTTACCAACCGCCCGCTTGTGAGGCGGTGACCGCTTGTGAGGCGGTGAGTGAACACC 4523
 Db 361 GTTACCAACCGCCCGCTTGTGAGGCGGTGACCGCTTGTGAGGCGGTGAGTGAACACC 420
 QY 4524 TCCGCGCTGACCGGATCCGCGCAACGAGCTGACCGCTGAGGCGCGGAGGCGGTGAGG 4583
 Db 421 TCCGCGCTGACCGGATCCGCGCAACGAGCTGACCGCTGAGGCGCGGAGGCGGTGAGG 480
 QY 4584 GCGTCGAAAGAGCTTGAACCTGATGAGCGCGGATCTTTCGCGAGCGCTTCCACAGT 4643
 Db 481 GCGTCGAAAGAGCTTGAACCTGATGAGCGCGGATCTTTCGCGAGCGCTTCCACAGT 540
 QY 4644 GCTGCTGCGGCGGCAAGGTGATGCGGCAAGTGCAGCGCGCTGCGGCGGTGACGAGCGG 4703
 Db 541 GCGGCGCGGCGGCAAGGTGATGCGGCAAGTGCAGCGCGCTGCGGCGGTGACGAGCGG 600
 QY 4704 GCGACGCTGCTCGGCGGTGACCAAGCGGCTGCGGCGGATCTTTCGCGAGGACCGGCTG 4763
 Db 601 GCGACGCTGCTCGGCGGTGACCAAGCGGCTGCGGCGGATCTTTCGCGAGGACCGGCTG 660
 QY 4764 AGTCCGAGCGAGCTGCGCGCGGTGAGCGGCGGATCTTTCGCGAGGACCGGCTGACT 4823
 Db 661 AGTCCGAGCGAGCTGCGCGCGGTGAGCGGCGGATCTTTCGCGAGGACCGGCTGACT 720
 QY 4824 TCAACCGGTGACGCTGACCGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 4883
 Db 721 TCAACCGGTGACGCTGACCGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 780
 QY 4884 GCGATGCGAGCTGCGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTGAGGCGGTG 4943
 Db 781 GCGATGCGAGCTGCGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTGAGGCGGTG 840
 QY 4944 GCGCGCGCGGTGAGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTGAGGCGGTG 5003
 Db 841 GCGCGCGCGGTGAGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTGAGGCGGTG 900
 QY 5004 TCGCGGTGCGGATCTGCTGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 5063
 Db 901 TCGCGGTGCGGATCTGCTGCGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 960
 QY 5064 CCGAAGATGCGCGGATCTGCTGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 5099
 Db 961 CCGAAGATGCGCGGATCTGCTGCGGATCTTTCGCGAGGACCGGCTGAGGCGGTG 996

RESULT 15

US-09-758-759-30/c
 ; Sequence 30, Application US/09758759
 ; Patent No. 6861513
 ; GENERAL INFORMATION:
 ; APPLICANT: Hosted, Thomas J.
 ; APPLICANT: Wang, Tim X.
 ; APPLICANT: Horan, Ann C.
 ; TITLE OF INVENTION: Evernimicin Biosynthetic Genes
 ; FILE REFERENCE: ID0983K US
 ; CURRENT APPLICATION NUMBER: US/09/758,759
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: US 60/175,751
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 30
 ; LENGTH: 1452
 ; TYPE: DNA

ORGANISM: Micromonospora carbonacea

; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1452)
 ; OTHER INFORMATION: evic
 US-09-758-759-30

Query Match 5.2%; Score 461.2; DB 3; Length 1452;
 Best Local Similarity 60.8%; Pred. No. 7.1e-67;
 Matches 820; Conservative 0; Mismatches 498; Indels 30; Gaps 3;

QY 5115 AGACGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5174
 Db 1450 AGACGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1391
 QY 5175 CGACGATGCTGATCTGCGCGGAGGATCAACCGGTGAGCGGAGGACCGGCTGCTG 5234
 Db 1390 GAGACAGGCGCTCACTGCGGAGCGGAGGATCAACCGGTGAGCGGAGGACCGGCTGCTG 1331
 QY 5235 CGAAGTCCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5294
 Db 1330 GGAAGTCTGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1271
 QY 5295 CCTCTCGAGAGTGCAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5354
 Db 1270 CCTCTCGAGAGTGCAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1211
 QY 5355 CGAAGTAGTCTGCGGAGGATCTTTCGCGGAGCGGAGGATCTTTCGCGGAGGATCTTTCG 5414
 Db 1210 CGAAGTAGTCTGCGGAGGATCTTTCGCGGAGCGGAGGATCTTTCGCGGAGGATCTTTCG 1163
 QY 5415 CGGTGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5474
 Db 1162 CGGTGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1103
 QY 5475 CGTGAAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5534
 Db 1102 CGTGAAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1043
 QY 5535 AGACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 5594
 Db 1042 AGACGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 983
 QY 5595 AGATGCTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5654
 Db 982 CCAAGGAGGAGTCCGCGCGGATCTTTCGCGGAGCGGAGGATCTTTCGCGGAGGATCTTTCGCGG 923
 QY 5655 TCGCGTGAATGTCGCGCAACCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 5714
 Db 922 GCAAGCTTCTGAGCGGAGCGGAGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 869
 QY 5715 CCGAAGTCCAGCTCAGCAGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGG 5774
 Db 868 CCGAAGTCCAGCTCAGCAGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGG 809
 QY 5775 AACGAGCAGTCCGCGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 5834
 Db 808 GCAAGTCCAGCTCAGCAGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGG 761
 QY 5835 CCGCGGCTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5894
 Db 760 CCGCGGCTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
 QY 5895 CGTGTCTGAGCGGAGCATCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGG 5954
 Db 700 GCGCGCTCAACCGGAGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGG 641
 QY 5955 GCTCGGCGGAGCTCTGAGGAGCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 6014
 Db 640 CCGCGGCGGAGCTCTGAGGAGCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 581
 QY 6015 AGGAACCTGCTCGGAGCTGAGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCGCGGAGGATCTTTCG 6074

```
Db      580 AGGAGCCCTGCTCCGACTGCGAGGACGTCAACCAAGCACCCGACCCCGCGCGGTGA 521
QY      6075 AGTACTCGATGAACTGGAACGCGGAAATCCGTTGAGAACCTCATGTAGTTGCTGCGGTTG 6134
Db      520 AGTAGTCGAGGTACTTGTGTCGCCGCCGACCGGTGCAACCGGTTAGTTGCTCGGGTTGG 461
QY      6135 CCTGCAGGGTTCGGCGAGAGCTGGAACGGCGTTGATGTTGCGAAGTTTCGGCCTTGAC 6194
Db      460 CTTGCAACCGTCCGGGAGAGCTGGCATGACGTTGATGTTGCCGGGCTTCACCTTGGCTTGA 401
QY      6195 CCAAGCGGTGCAAGTACGCGCTCGAACTCCGCGACGATCAGACCGAGGAACCCGATCTCG 6254
Db      400 GCAAGCAGTACGGTGTCCCGTCGACGACTTGACGAGCATGCCCAGGATGCGGATCTCCG 341
QY      6255 GTTGGAGATGATGAGTTGATTCAGTCCCGTCGCCATCCGAAGTTGTCGAGCGTGA 6314
Db      340 GCTGTTGATGATGAGCTGCTGTGTCATTCGCGCACCGCGCGTAGTGTGTCGACGTGA 281
QY      6315 GGCCTCGATGAGAGAAAGCGCCGAGTCTGTGCGCCAGCCGACCGTCTCCGGGTGA 6374
Db      280 GCCCTCGATCAGAAAGAACCGGCGCGCTCTGTGTCGAGGTTGCCGGTCAACGGGTGA 221
QY      6375 ACGACCAAGCTTCCATGCTGTGAAGGCACTCGGTGCACTCGAGCCGATGCTCGGCG 6434
Db      220 ACGCCCAACCGGCGACCGGTCCAGCGGCAACCGGTCCACCCGCGCAGTAGGTGACCGGG 161
QY      6435 CGCGTGGGCCAACGATCGTGAAGTTC 6462
Db      160 TCCGCTCGGCGAACCAAGAGAGAAAGTC 133
```

Search completed: April 7, 2006, 02:59:03
Job time : 1490.43 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 01:33:46 ; Search time 5144.9 Seconds
(without alignments)
11689.676 Million cell updates/sec

Title: US-10-611-442-1

Perfect score: 8942

Sequence: 1 ggcgcgtcgcacccatcgcgtg.....gtggagccacggagagatcc 9024

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N_Geneseq_21:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2002s:*
6: geneseqn2003s:*
7: geneseqn2004s:*
8: geneseqn2005s:*
9: geneseqn2006s:*
10: geneseqn2007s:*
11: geneseqn2008s:*
12: geneseqn2009s:*
13: geneseqn2010s:*
14: geneseqn2011s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8942	100.0	9024	12	AD114147
2	3037.4	34.0	8160	2	AAZ25774
3	3032.6	33.9	8051	2	AAZ72685
4	2606.8	29.2	47981	4	AAZ30757
5	1309.4	14.6	12441	3	AAZ87284
6	1309.4	14.6	13613	6	AAZ39043
7	1309.4	14.6	13613	12	ADL91931
8	1307.8	14.6	13613	3	AAZ87319
9	1241.2	13.9	60196	10	AAZ55810
10	1203.2	13.5	84428	12	ADL15447
11	729.4	8.2	1565	3	AAZ55638
12	729.4	8.2	1565	3	AAZ56006
13	729.4	8.2	1565	8	ADL09423
14	729.4	8.2	1565	10	ADL15447
15	729.4	8.2	1565	10	ADL15447
16	722	8.1	1458	6	AAZ39045
17	722	8.1	1458	12	ADL15447
18	720.6	8.1	85915	13	ADL15447
19	720.6	8.1	85915	13	ADL15447

C	20	711.6	8.0	3412	2	AAZ25772	AAZ25772 S. erythra
C	21	701.4	7.8	3756	2	AAZ72684	AAZ72684 Sugar bio
C	22	657.2	7.5	1467	10	AAZ55828	AAZ55828 Micromono
C	23	612.8	6.9	17596	12	ADL14148	ADL14148 M. megalio
C	24	596.6	6.7	1227	10	AAZ55827	AAZ55827 Micromono
C	25	595.4	6.7	1248	3	AAZ87286	AAZ87286 S. venez
C	26	595.4	6.7	1248	6	AAZ39044	AAZ39044 Streptomy
C	27	595.4	6.7	1248	12	ADL91935	ADL91935 Streptomy
C	28	519.2	5.8	5970	3	AAZ75635	AAZ75635 Nucleotid
C	29	519.2	5.8	5970	3	AAZ56003	AAZ56003 Contig 00
C	30	519.2	5.8	5970	8	ADL09420	ADL09420 S. venez
C	31	519.2	5.8	5970	10	ADL15447	ADL15447 S. venez
C	32	519.2	5.8	5970	10	AAZ56002	AAZ56002 Contig 00
C	33	500.4	5.6	45055	10	AAZ66808	AAZ66808 Orthosomy
C	34	496.6	5.3	1401	10	AAZ66809	AAZ66809 Streptomy
C	35	474.2	5.3	1404	8	AAZ37554	AAZ37554 Streptomy
C	36	474.2	5.3	59816	8	AAZ37515	AAZ37515 Streptomy
C	37	474.2	5.3	59816	8	AAZ37515	AAZ37515 Streptomy
C	38	471.2	5.3	4461	14	AAZ48864	AAZ48864 Angolamyc
C	39	463.4	5.2	30943	12	ADL97550	ADL97550 S. ambofac
C	40	463.4	5.2	30943	12	ADL97550	ADL97550 S. ambofac
C	41	461.2	5.2	1455	10	AAZ66810	AAZ66810 Orthosomy
C	42	461.2	5.2	37116	10	AAZ66810	AAZ66810 Orthosomy
C	43	461.2	5.2	109519	5	AAZ08693	AAZ08693 Micromono
C	44	451.2	5.0	1410	12	ADL97559	ADL97559 S. ambofac
C	45	451.2	5.0	1410	12	ADL97559	ADL97559 S. ambofac

ALIGNMENTS

RESULT 1
AD114147
ID AD114147 standard; DNA; 9024 BP.
XX
XX AD114147;
AC
XX
DT 15-APR-2004 (first entry)
DE M. megalomicea cosmid PKOS079-138B SEQ ID NO:1.
XX
XX ds: polyketide; enzyme; MegR; MegF; MegG; MegK; MegCIV; MegCV; MegBVI; MegBIII;
KM MegL; MegM; megosamine; megalomycin.
XX
OS Micromonospora megalomicea.
XX
XX WO2004003169-A2.
PN
PD 08-JAN-2004.
XX
XX 30-JUN-2003; 2003WO-US020681.
XX
XX 28-JUN-2002; 2002US-0393016P.
PR
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H;
XX WPI; 2004-203379/19.
XX
XX Novel isolated, purified, or recombinant nucleic acid comprising
PT polyketide modifying gene, there gene encodes polyketide modifying enzyme
PT e.g., MegR, MegK, or MegM enzymes useful for producing modified
PT polyketide.
XX
XX Example 2; SEQ ID NO 1; 51pp; English.
XX
XX The invention relates to a novel isolated, purified, or recombinant
XX nucleic acid (I) comprising a polyketide modifying gene, where the gene
XX encodes a polyketide modifying enzyme chosen from MegR, MegF, MegG,
XX MegCIV, MegCV, MegBVI, MegBIII, MegL, and MegM enzymes. A method of the
XX invention is useful for producing a modified polyketide, which involves
XX culturing a recombinant cell comprising the recombinant nucleic acid

under conditions in which the cell expresses a product of a gene encoded by the nucleic acid under conditions in which the unmodified polypeptide is present, and producing the modified polypeptide. The cell produces megasamine and can attach megasamine to a polypeptide, where the cell, it is naturally occurring non-recombinant state cannot produce megasamine. The present sequence contains upstream megalomycin modification enzyme genes.

Sequence 9024 BP; 1324 A; 3022 C; 3127 G; 1469 T; 0 U; 82 Other;

Query Match 100.0%; Score 8942; DB 12; Length 9024;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9024; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GCGGCTTCGATTCACCATGATCGCTTAATGTCGGGTCCATGTCCTTCGATGGGGAT 60
DB 1 GCGGCTTCGATTCACCATGATCGCTTAATGTCGGGTCCATGTCCTTCGATGGGGAT 60
QY 61 GTAGTCATTTGCAAAATCCGAGACCGTGTAAAGCTTCGAGTCCGCTGTTTC 120
DB 61 GTAGTCATTTGCAAAATCCGAGACCGTGTAAAGCTTCGAGTCCGCTGTTTC 120
QY 121 GTCAATTCCTTCATGCGCGGAAAGACGCGCAAGCGCAAGAAATGTTCTCGCA 180
DB 121 GTCAATTCCTTCATGCGCGGAAAGACGCGCAAGCGCAAGAAATGTTCTCGCA 180
QY 181 ATGCTCCCTGTCACGCGGACCAAGGTGTTCCGGTCTCGTCTCGTCTCGAGCTCGG 240
DB 181 ATGCTCCCTGTCACGCGGACCAAGGTGTTCCGGTCTCGTCTCGTCTCGAGCTCGG 240
QY 241 GACGACGAGCGCGCGTCAAGAGGCTCAACACCTTCAGACGTAATTCATCTGCGC 300
DB 241 GACGACGAGCGCGCGTCAAGAGGCTCAACACCTTCAGACGTAATTCATCTGCGC 300
QY 301 AAGATGTTCTGCGCGGTCAACCGGTCTGCGCGCGGAGGTCAACCGGAGTCTGCTATC 360
DB 301 AAGATGTTCTGCGCGGTCAACCGGTCTGCGCGCGGAGGTCAACCGGAGTCTGCTATC 360
QY 361 ACCGCGCGCGCGCTATCTGCTCGCGGTGACCGGATCGCGCTCGACGTCGCGGAGTAC 420
DB 361 ACCGCGCGCGCGCTATCTGCTCGCGGTGACCGGATCGCGCTCGACGTCGCGGAGTAC 420
QY 421 CAGCGTCAATGTCGCGCGGCTGCGCGCGCTCGCGCTCGCGCAAGCTGACAGGCAAC 480
DB 421 CAGCGTCAATGTCGCGCGGCTGCGCGCGCTCGCGCTCGCGCAAGCTGACAGGCAAC 480
QY 481 CGCAGACTCAACGAGCGCTGCGCTCTGCGCGCGGCGCGCGCTGTCGACGTCGCCGTG 540
DB 481 CGCAGACTCAACGAGCGCTGCGCTCTGCGCGCGGCGCGCGCTGTCGACGTCGCCGTG 540
QY 541 GGTGCGGTGTCGAGATCGAAGGCTGCGCACTGAGGAGTCTGCTCATGCGCAGCA 600
DB 541 GGTGCGGTGTCGAGATCGAAGGCTGCGCACTGAGGAGTCTGCTCATGCGCAGCA 600
QY 601 TACCTGTGCGCGCGAAGTTGCTGACGAGGATGTACCGGAGGCGCTCATGCGAGTGAAC 660
DB 601 TACCTGTGCGCGCGAAGTTGCTGACGAGGATGTACCGGAGGCGCTCATGCGAGTGAAC 660
QY 661 GCGCTCACCGCGAGAAATCCGCTGCAAGAGGCTTCAGAGCGCAGTACATGCGGCGCTG 720
DB 661 GCGCTCACCGCGAGAAATCCGCTGCAAGAGGCTTCAGAGCGCAGTACATGCGGCGCTG 720
QY 721 CATCTCAATGTCGACGCGCGGAGGCTTTCACACCGGTGCTGCAACCTTC 780
DB 721 CATCTCAATGTCGACGCGCGGAGGCTTTCACACCGGTGCTGCAACCTTC 780
QY 781 GTTCAGAACTGGGTCTGGAACCGGAGCGCAGGTGCAACCGATCCACAGGCGATCTCTG 840
DB 781 GTTCAGAACTGGGTCTGGAACCGGAGCGCAGGTGCAACCGATCCACAGGCGATCTCTG 840
QY 841 AACGCGGAGACCGACTTCGAGAGACGATCTGCGCGTCACTCGTCCGTTCCGTCGAGGTC 900
DB 841 AACGCGGAGACCGACTTCGAGAGACGATCTGCGCGTCACTCGTCCGTTCCGTCGAGGTC 900
```

```

QY 901 GCCGCAAGATTGGGTGCGGTCCGGGTCCGGCGAGCTGACCGATTACCGCTACGCG 960
DB 901 GCCGCAAGATTGGGTGCGGTCCGGGTCCGGCGAGCTGACCGATTACCGCTACGCG 960
QY 961 GACACCTTGAGCGGACATCAACGACTTGGCGGAAATTCGACATCTGTGCGCGGGGGA 1020
DB 961 GACACCTTGAGCGGACATCAACGACTTGGCGGAAATTCGACATCTGTGCGCGGGGGA 1020
QY 1021 CGGGTTGAGCGAAGCGTGGGGAACCATGACACTATGAAACGATCCGAGCATG 1080
DB 1021 CGGGTTGAGCGAAGCGTGGGGAACCATGACACTATGAAACGATCCGAGCATG 1080
QY 1081 GCCGAGAGCGCGTTCTGCTGACCTGCGCGGTGATGCGCGACAGGACCCGCTGCG 1140
DB 1081 GCCGAGAGCGCGTTCTGCTGACCTGCGCGGTGATGCGCGACAGGACCCGCTGCG 1140
QY 1141 CAGGACCAATGACGGGTCTGACATCTTCGCGCAAGTGAAGTACCGGAGTCTCGC 1200
DB 1141 CAGGACCAATGACGGGTCTGACATCTTCGCGCAAGTGAAGTACCGGAGTCTCGC 1200
QY 1201 GACACCGGACCTTCTGCTCGGACCCGACCGGCTCATGAGGGGCGCGACCGCG 1260
DB 1201 GACACCGGACCTTCTGCTCGGACCCGACCGGCTCATGAGGGGCGCGACCGCG 1260
QY 1261 GGGATGATCCACGAGATCGACCCGCGAGGACCGGGCCCTGCGCAAGTCTGACAGC 1320
DB 1261 GGGATGATCCACGAGATCGACCCGCGAGGACCGGGCCCTGCGCAAGTCTGACAGC 1320
QY 1321 GCTTTCACCCCGGTGATGATGCGGACCTCGAACCGGACATCCGGAGGTACCCCGTGC 1380
DB 1321 GCTTTCACCCCGGTGATGATGCGGACCTCGAACCGGACATCCGGAGGTACCCCGTGC 1380
QY 1381 CTGCTGCGCGACGCGCGGTGACCGGCTGACCTGCTGAGGCGCTCGCTTCCGCTGCG 1440
DB 1381 CTGCTGCGCGACGCGCGGTGACCGGCTGACCTGCTGAGGCGCTCGCTTCCGCTGCG 1440
QY 1441 GTCAAGATCGTCCGCGAGCTGCTGCGGTGCGCGGATGACCAAGACATTCGATGAC 1500
DB 1441 GTCAAGATCGTCCGCGAGCTGCTGCGGTGCGCGGATGACCAAGACATTCGATGAC 1500
QY 1501 TGGTCCGCGCGCTGCTGATCACTCAATGAGACCAACCGGATCCGCGCTGATCGAA 1560
DB 1501 TGGTCCGCGCGCTGCTGATCACTCAATGAGACCAACCGGATCCGCGCTGATCGAA 1560
QY 1561 CGCATCAATGACGAGTGTGAACCGGCTGACCTGACCTGCTGACAGGTGTCGAGGACG 1620
DB 1561 CGCATCAATGACGAGTGTGAACCGGCTGACCTGACCTGCTGACAGGTGTCGAGGACG 1620
QY 1621 CGGGCGGACCCCGGAGACGACTGATCTCCGGCTGCTGCGCGAGGTGACCGGCGC 1680
DB 1621 CGGGCGGACCCCGGAGACGACTGATCTCCGGCTGCTGCGCGAGGTGACCGGCGC 1680
QY 1681 ACCCTCGACGAGTGAAGCGCGCACTTCTCAACAGGTGCTGCGCGGCACTTC 1740
DB 1681 ACCCTCGACGAGTGAAGCGCGCACTTCTCAACAGGTGCTGCGCGGCACTTC 1740
QY 1741 ACCACCAACGCTGCTGCGGGAACATGCTCGGACCCCTCGACGAGCAACCGGAGTAC 1800
DB 1741 ACCACCAACGCTGCTGCGGGAACATGCTCGGACCCCTCGACGAGCAACCGGAGTAC 1800
QY 1801 ACCGCGCGCGCGAGGACCCCGGCTGATGCGCGCGATCATGAGAGGTGTCGTTTC 1860
DB 1801 ACCGCGCGCGCGAGGACCCCGGCTGATGCGCGCGATCATGAGAGGTGTCGTTTC 1860
QY 1861 CGCCCGCGCTTCCCGAGATGACGCAACGACGAGGCGCACCGCTGCTGCTGCGGTC 1920
DB 1861 CGCCCGCGCTTCCCGAGATGACGCAACGACGAGGCGCACCGCTGCTGCTGCGGTC 1920
QY 1921 GAGATCCCGGCGAGTGAATGCTGCAACCTGAGGCTCTGCGGCAACCGGATCCCTTC 1980
DB 1921 GAGATCCCGGCGAGTGAATGCTGCAACCTGAGGCTCTGCGGCAACCGGATCCCTTC 1980
```

QY 1981 GCGCATCCGACCCGGACAGCTTTCGACCCGCTCCGCGAAGATCGGTGTGCCGCGAGCTC 2040
DB 1981 GCGCATCCGACCCGGACAGCTTTCGACCCGCTCCGCGAAGATCGGTGTGCCGCGAGCTC 2040
QY 2041 TCCCTTCGGGACCGCGGTGCACTTCTGTCTCGGTGCCCCGCTGGGGCGCTTGAAGACAG 2100
DB 2041 TCCCTTCGGGACCGCGGTGCACTTCTGTCTCGGTGCCCCGCTGGGGCGCTTGAAGACAG 2100
QY 2101 GTCGCCCTGAGAGATCATCGCCGGGTACGGTGCAGCTGCGCGTGCAGCCGGACGACGAC 2160
DB 2101 GTCGCCCTGAGAGATCATCGCCGGGTACGGTGCAGCTGCGCGTGCAGCCGGACGACGAC 2160
QY 2161 ACCTGCTGTCATTTCGACAGATCGTCTCGGCAACCGGCACTTCCGGTCTGAGCGCG 2220
DB 2161 ACCTGCTGTCATTTCGACAGATCGTCTCGGCAACCGGCACTTCCGGTCTGAGCGCG 2220
QY 2221 GTCACCCCGGCGGAGTCCGCTGAAACCTTGCCTCGACGCGCGGNNNNNNNNNN 2280
DB 2221 GTCACCCCGGCGGAGTCCGCTGAAACCTTGCCTCGACGCGCGGNNNNNNNNNN 2280
QY 2281 NNN 2340
DB 2281 NNN 2340
QY 2341 NNN 2400
DB 2341 NNN 2400
QY 2401 CAGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2460
DB 2401 CAGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2460
QY 2461 GGTGACCGGACGTGTGCAAAACGTGAGAGAGAACTCCCTCCCGCGCTGCGGATGATGCT 2520
DB 2461 GGTGACCGGACGTGTGCAAAACGTGAGAGAGAACTCCCTCCCGCGCTGCGGATGATGCT 2520
QY 2521 GCGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2580
DB 2521 GCGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2580
QY 2581 GCGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2640
DB 2581 GCGGCCCCGATGCTGATCCCAACCCGTCGGCGAGCTCCGTTGAGTTGTTGAGCCGCG 2640
QY 2641 GAGCTTCGAGAGGTGCACTGCGACCGCACTTGGGATGGGCGGTCCGAGTCAAGTCA 2700
DB 2641 GAGCTTCGAGAGGTGCACTGCGACCGCACTTGGGATGGGCGGTCCGAGTCAAGTCA 2700
QY 2701 TTTCCGCGCGGATCCGCAACAGCTGGGCGTGCACCCCGACGAGGCTTTCGAACGCGTA 2760
DB 2701 TTTCCGCGCGGATCCGCAACAGCTGGGCGTGCACCCCGACGAGGCTTTCGAACGCGTA 2760
QY 2761 ACCGAGGTGCTGCACTGCGACCGCACTTGGGATGGGCGGTCCGAGTCAAGTCA 2820
DB 2761 ACCGAGGTGCTGCACTGCGACCGCACTTGGGATGGGCGGTCCGAGTCAAGTCA 2820
QY 2821 GCGCTCCCGGAACTCGGTTCGCTTCGCTCGGCGGCGTGAAGTCAAGTCAAGTCA 2880
DB 2821 GCGCTCCCGGAACTCGGTTCGCTTCGCTCGGCGGCGTGAAGTCAAGTCAAGTCA 2880
QY 2881 GTCCTCGCGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2940
DB 2881 GTCCTCGCGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2940
QY 2941 GATGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3000
DB 2941 GATGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3000
QY 3001 GCGGAGCGGAGCGGTGCGCGCTTTCGCGCGGATCCGAGAGAGTCCGAGAGTCTT 3060
DB 3001 GCGGAGCGGAGCGGTGCGCGCTTTCGCGCGGATCCGAGAGAGTCCGAGAGTCTT 3060
QY 3061 CTTGACCGGTTGAAAGCGGACGCTTTCGCGGTGCGGTGAGTCTGCTGTTGCTGAG 3120

DB 3061 CTTGACCGGTTGAAAGCGGACGCTTTCGCGGTGCGGTGAGTCTGCTGTTGCTGAG 3120
QY 3121 CCCGTAAGAGGATACGACCCGCGGAGGTCCCAAGCCGCGGCTGAGCGGCGAGCT 3180
DB 3121 CCCGTAAGAGGATACGACCCGCGGAGGTCCCAAGCCGCGGCTGAGCGGCGAGCT 3180
QY 3181 CTGTTGAGTGAAGGCGGAAAGCTTGTGTGAAGTGAAGGCGGAAAGCCCGCGGCGG 3240
DB 3181 CTGTTGAGTGAAGGCGGAAAGCTTGTGTGAAGTGAAGGCGGAAAGCCCGCGGCGG 3240
QY 3241 GTGCGACCAAGCGCCCGGATACCCGCGGTTGTCAAGGTTCCAGGCGCGGAGAGTGA 3300
DB 3241 GTGCGACCAAGCGCCCGGATACCCGCGGTTGTGAAGTTCAGGCTCCGAGAGAGTGA 3300
QY 3301 CATGCGGTGAGGTTTTCGCAAGGCACTTCGTGATGACCGAGGTGAGATCTTCTGTTCC 3360
DB 3301 CATGCGGTGAGGTTTTCGCAAGGCACTTCGTGATGACCGAGGTGAGATCTTCTGTTCC 3360
QY 3361 GGGGACCAAGCGCTCCGCGGTGTAATTGAGCACTGTAACCGGTAAGAGTGAAGCG 3420
DB 3361 GGGGACCAAGCGCTCCGCGGTGTAATTGAGCACTGTAACCGGTAAGAGTGAAGCG 3420
QY 3421 GAAATGCAAGTCCGCGCGGAGTGAAGCCCGGAGTCCCAAGGAGGCGGAGTCCGCG 3480
DB 3421 GAAATGCAAGTCCGCGCGGAGTGAAGCCCGGAGTCCCAAGGAGGCGGAGTCCGCG 3480
QY 3481 TACCGAGCTTCAAGCGGCGGCGGTTCAAGCGGCTGATGTTGCTCAAGTCTT 3540
DB 3481 TACCGAGCTTCAAGCGGCGGCGGTTCAAGCGGCTGATGTTGCTCAAGTCTT 3540
QY 3541 GCGGCGGCGGCGGCTTCCAGCGGCGGCGGCTTCCAGTGGGAGTGCAGTCAAGTCAAG 3600
DB 3541 GCGGCGGCGGCGGCTTCCAGCGGCGGCGGCTTCCAGTGGGAGTGCAGTCAAGTCAAG 3600
QY 3601 CAGCTGCGCAAGCCCGGCGGCTTCCAGGTCAGAGTCAAGCGGAGTCTTCCAGGAGT 3660
DB 3601 CAGCTGCGCAAGCCCGGCGGCTTCCAGGTCAGAGTCAAGCGGAGTCTTCCAGGAGT 3660
QY 3661 GAAAGGAGGCGGCGGATGAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3720
DB 3661 GAAAGGAGGCGGCGGATGAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3720
QY 3721 GGTGCGGAGTCCGCAACCGGTCAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3780
DB 3721 GGTGCGGAGTCCGCAACCGGTCAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3780
QY 3781 CCGCGGAGTCCGCAACCGGTCAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3840
DB 3781 CCGCGGAGTCCGCAACCGGTCAGGCGGAGTTCGAGCGGAGCGGCTTCTGTTGCG 3840
QY 3841 AGGAGTGTCTTATGACGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3900
DB 3841 AGGAGTGTCTTATGACGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 3900
QY 3901 CGTTGCGTTTCCGCTTGTCCCACTTCGCGTGAATCAAGCAAGTCAAGTCAAGTCA 3960
DB 3901 CGTTGCGTTTCCGCTTGTCCCACTTCGCGTGAATCAAGCAAGTCAAGTCAAGTCA 3960
QY 3961 GATGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4020
DB 3961 GATGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4020
QY 4021 CGAGCAACCGGTTGCGGAGGCACTTCGCTGATGACCGGAGGAGTCCGAGTCCG 4080
DB 4021 CGAGCAACCGGTTGCGGAGGCACTTCGCTGATGACCGGAGGAGTCCGAGTCCG 4080
QY 4081 AGGAGTGTCTTATGACGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4140
DB 4081 AGGAGTGTCTTATGACGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4140
QY 4141 GCAGAGTCCGTCAGACCGGTCAGGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 4200

D	b	4141	GCAGAAATCGCTGACGCCCGGTACACGGTTCGATGCCCGGTGGCGGTGCCTTCCACACAGATGA	4200
Q	y	4201	TCACGTACTGCTAGTTGCTCTCTCTGTCGGGCGGAAACGAGTGCACGGTGAACCCCGGTGA	4260
D	b	4201	TCACGTACTGCTAGTTGCTCTCTCTGTCGGGCGGAAACGAGTGCACGGTGAACCCCGGTGA	4260
Q	y	4261	CGTGGCCAGCTCGCTGTGTGTAAAGACCGGTGTGTAAACCCGGTGTGTCTCCCTGTCTCGG	4320
D	b	4261	CGTGGCCAGCTCGCTGTGTGTAAAGACCGGTGTGTAAACCCGGTGTGTCTCCCTGTCTCGG	4320
Q	y	4321	CGAACCGGTGCAGGGAAGGTGAGCCCCCATGGCCGCGGCGCACTGCATCTTGGCCGTTGG	4380
D	b	4321	CGAACCGGTGCAGGGAAGGTGAGCCCCCATGGCCGCGGCGCACTGCATCTTGGCCGTTGG	4380
Q	y	4381	TGCCGATCTCGGTGACCACTTGTGTCGGGCGGATGCGGAAGTTGTGATGGCCCGATCC	4440
D	b	4381	TGCCGATCTCGGTGACCACTTGTGTCGGGCGGATGCGGAAGTTGTGATGGCCCGATCC	4440
Q	y	4441	GTTTCGGCCAGACGAGCCGTGTCGTGGTGAACACCGCCCCGCTCTGAAAGCGGTGACCGCTT	4500
D	b	4441	GTTTCGGCCAGACGAGCCGTGTCGTGGTGAACACCGCCCCGCTCTGAAAGCGGTGACCGCTT	4500
Q	y	4501	TGCTGGGGGTGGAACCTGAAACAACCTCGGGGTGACAGGATCCGGCCACCGGAACGTTCACCG	4560
D	b	4501	TGCTGGGGGTGGAACCTGAAACAACCTCGGGGTGACAGGATCCGGCCACCGGAACGTTCACCG	4560
Q	y	4561	TCGTGACAGCCCAAGGCGCGTGGGCGGCGGTGGAAGAAAGACTTGAACCTGTGTGTGGCGGGA	4620
D	b	4561	TCGTGACAGCCCAAGGCGCGTGGGCGGCGGTGGAAGAAAGACTTGAACCTGTGTGTGGCGGGA	4620
Q	y	4621	TCTTTCGCCAGCGCTTCCACAGGTGTGTGTGGGCCCCACAGTGTATCCGCCATCTGCCTG	4680
D	b	4621	TCTTTCGCCAGCGCTTCCACAGGTGTGTGTGGGCCCCACAGTGTATCCGCCATCTGCCTG	4680
Q	y	4681	CGGTCTGGGGGGTGTACAGAGCGGGCGGACGTGTGTCCGGGTGCACAGACCGGTGC	4740
D	b	4681	CGGTCTGGGGGGTGTACAGAGCGGGCGGACGTGTGTCCGGGTGCACAGACCGGTGC	4740
Q	y	4741	CGACGTGCACAACAACCGGTGTGAGTCCGACCAAGCTTCGCGCGTGGCGGTGGCCGGA	4800
D	b	4741	CGACGTGCACAACAACCGGTGTGAGTCCGACCAAGCTTCGCGCGTGGCGGTGGCCGGA	4800
Q	y	4801	AGGTCAATTGACGGCATGATCACTTCAACCGGTGACGTCAACCGGCCCGGACACAGTTTCCA	4860
D	b	4801	AGGTCAATTGACGGCATGATCACTTCAACCGGTGACGTCAACCGGCCCGGACACAGTTTCCA	4860
Q	y	4861	GGGCGACCGGTGGCGTTGGAGGTGGCGATGACGTGCCCTGATCCCGACACAGTCCGCGAACCC	4920
D	b	4861	GGGCGACCGGTGGCGTTGGAGGTGGCGATGACGTGCCCTGATCCCGACACAGTCCGCGAACCC	4920
Q	y	4921	GGGCGCTCGAATCTCCGCGACACGAGGGCCCGCGTGTGTGAGCGCAATGTTGTTCAAGGCC	4980
D	b	4921	GGGCGCTCGAATCTCCGCGACACGAGGGCCCGCGTGTGTGAGCGCAATGTTGTTCAAGGCC	4980
Q	y	4981	ACTCCAGGCGGGCGAGAAACCGCTGCGCGGTGCTGCGCGGCGCCACGTGCAGGG	5040
D	b	4981	ACTCCAGGCGGGCGAGAAACCGCTGCGCGGTGCTGCGCGGCGCCACGTGCAGGG	5040
Q	y	5041	GGTGACGGAACCGCTCTCGGACCGCGGAAGATCGCCAGATCGGTGTGTATCAGGCTTCAACGC	5100
D	b	5041	GGTGACGGAACCGCTCTCGGACCGCGGAAGATCGCCAGATCGGTGTGTATCAGGCTTCAACGC	5100
Q	y	5101	CGTGGCCCGGTGTGTGAACGCGGACACGCGACGGCAAGGCTGGCGACGTGGATGTTGATC	5160
D	b	5101	CGTGGCCCGGTGTGTGAACGCGGACACGCGACGGCAAGGCTGGCGACGTGGATGTTGATC	5160
Q	y	5161	GTAAGTGGCTGTGCGCGAGCAATTCCGATGATCTGGCCGAAGTCAATCCACCGGTGTCCGG	5220
D	b	5161	GTAAGTGGCTGTGCGCGAGCAATTCCGATGATCTGGCCGAAGTCAATCCACCGGTGTCCGG	5220
Q	y	5221	GGCACCGCGCTGTGTGAAGTCCGCGGGCACTCTGACGACATGTATCCGATTTCTGTTTGG	5280
D	b	5221	GGCACCGCGCTGTGTGAAGTCCGCGGGCACTCTGACGACATGTATCCGATTTCTGTTTGG	5280

QY	5281	GTAGAAACCGCCACCCCTCTCTCGAAGTGCACGGCCGTCGATAGGAGATCTGGGATACCGCGGAC	5340
Db	5281	GTAGAAACCGCCACCCCTCTCTCGAAGTGCACGGCCGTCGATAGGAGATCTGGGATACCGCGGAC	5340
QY	5341	GTCACAGACGTAAGTCAGAGGTAGAGTCCGGCGGGATTCTCCGAGCAGTCGATGATGTTGTC	5400
Db	5341	GTCACAGACGTAAGTCAGAGGTAGAGTCCGGCGGGATTCTCTCGAGCAGTCGATGATGTTGTC	5400
QY	5401	CGGCTGACAGTGCACCGTGCAGCCGCACTCGGCGACGTCGTGGCCACCCGCTTCGGTGC	5460
Db	5401	CGGCTGACAGTGCACCGTGCAGCCGCCCACTCGGCGACGTCGTGGCCACCCGCTTCGGTGC	5460
QY	5461	CAGGTGCACGAGCGCTGAGGGATGCGTCGATCTCTTGACAGAGAGGGGAGACATAGCC	5520
Db	5461	CAGGTGCACGAGCGCTGAGGGATGCGTCGATCTCTTGACAGAGAGGGGAGACATAGCC	5520
QY	5521	GTGGTTGGCGGGGGAGAGAGCGGTTGCATCCAGGACCTCACCTGCGGGTGACTAGTGTGT	5580
Db	5521	GTGGTTGGCGGGGGAGAGAGCGGTTGCATCCAGGACCTCACCTGCGGGTGACTAGTGTGT	5580
QY	5581	CACGAGACGCGCCGAGATGCTGAAGTACTTCTGTCTCTGTGTTGATGCGTCGTCCGC	5640
Db	5581	CACGAGAGCGCCGAGATGCTGAAGTACTTCTGTCTCTGTGTTGATGCGTCGTCCGC	5640
QY	5641	CCGGATTCACCCGCTCGGTGCATGTCCGCGCAACCGCGGGGCGCTTGAAGAACTCACG	5700
Db	5641	CCGGATTCACCCGCTCGGTGCATGTCCGCGCAACCGCGGGGCGCTTGAAGAACTCACG	5700
QY	5701	TGTGGTGCAGTACGTCCGAGATCCAGCTCAGCAGAGGGGTTCATGTTGTGAACCGAGGTGCC	5760
Db	5701	TGTGGTGCAGTACGTCCGAGATCCAGCTCAGCAGAGGGGTTCATGTTGTGAACCGAGGTGCC	5760
QY	5761	CGCCCGCACAGAGAAACGACAGCTCGGGTCTTCAGAGAACCCCTCGAGGCGAGTCGTCAG	5820
Db	5761	CGCCCGCACAGAGAAACGACAGCTCGGGTCTTCAGAGAGACCCCTCGAGGCGAGTCGTCAG	5820
QY	5821	TACCCGAGCGACGGGCCCCCGGCTTCCACCTCGTCCGATGATGACGGCCGACAGCCGAG	5880
Db	5821	TACCCGAGCGACGGGCCCCCGGCTTCCACCTCGTCCGATGATGACGGCCGACAGCCGAG	5880
QY	5881	GTCCATGTTTCAACAAGTGTGTCGTAAGCGAGCAATCCCGCGACGTGGGGAGGGGTGAGCA	5940
Db	5881	GTCCATGTTTCAACAAGTGTGTCGTAAGCGAGCAATCCCGCGACGTGGGGAGGGGTGAGCA	5940
QY	5941	CCTGCAAGTTCGGGGTCTCGGGCGGGGCTCTCGAAGACCTGCACACCATGTGGCGTTGCG	6000
Db	5941	CCTGCAAGTTCGGGGTCTCGGGCGGGGCTCTCGAAGACCTGCACACCATGTGGCGTTGCG	6000
QY	6001	TTTTCGCGAGAACCAAGAAACCTTGCTCGGACTTGAGAGAGCTGCACGAGGATCCGGTGGGG	6060
Db	6001	TTTTCGCGAGAAACCAAGAAACCTTGCTCGGACTTGAGAGAGCTGCACGAGGATCCGGTGGGG	6060
QY	6061	GCGGGTCCCGTCCGAAGTATCTCATGAACTGGAACGGGGATCCGTTGTGGAACCTTCATGTA	6120
Db	6061	GCGGGTCCCGTCCGAAGTATCTCATGAACTGGAACGGGGATCCGTTGTGGAACCTTCATGTA	6120
QY	6121	GTTGCTGAGGGTGGCCGACGAGGTCCGCGCGAGAGCTGGAACGGCGTGTGATGTGGCCAGGTTCC	6180
Db	6121	GTTGCTGAGGGTGGCCGACGAGGTCCGCGCGAGAGCTGGAACGGCGGTGATGTGGCCAGGTTCC	6180
QY	6181	GACCTTGAGCTGCACAGACGTCGAGTACGCGCTGGAATTCGCGCAGCATTCAGACCGAG	6240
Db	6181	GACCTTGAGCTGCACAGACGTCGAGTACGCGCTGGAATTCGCGCAGCATTCAGACCGAG	6240
QY	6241	GAACCCGATCTCGGGTTGACGATGATGGGTTGAATCCAGTCCGCTCCGATCCGAAATT	6300
Db	6241	GAACCCGATCTCGGGTTGACGATGATGGGTTGAATCCAGTCCGCTCCGATCCGAAATT	6300
QY	6301	GGTCCGACAGTGCACGCGCTTCGATGGAAGAAAGAGCGCCGAGAGTGGTGGCCAGCGCAAC	6360
Db	6301	GGTCCGACAGTGCACGCGCTTCGATGGAAGAAAGAGCGCCGAGAGTGGTGGCCAGCGCAAC	6360

QY 6361 GTCTCCGGGTGGAACGACGCTTCATGTCGTGAAGGCACTCGTGCACCTCGAG 6420
Db 6361 GTCTCCGGGTGGAACGACGCTTCATGTCGTGAAGGCACTCGTGCACCTCGAG 6420
QY 6421 CCGATGCTCGGCGCGCGGTGAGCCACCACTCGTGAAGCTGTCGTGCGGTGAGAG 6480
Db 6421 CCGATGCTCGGCGCGCGGTGAGCCACCACTCGTGAAGCTGTCGTGCGGTGAGAG 6480
QY 6481 TCCGCGGTGCGGATCAGGAAACGATTTGATGAGATTCCGGAAGTCGATGACC 6540
Db 6481 TCCGCGGTGCGGATCAGGAAACGATTTGATGAGATTCCGGAAGTCGATGACC 6540
QY 6541 GTTACCCGATCCCGCAATCGCTCTCCCGTATGTCGTGCGCGGTGAGAGAG 6600
Db 6541 GTTACCCGATCCCGCAATCGCTCTCCCGTATGTCGTGCGCGGTGAGAGAG 6600
QY 6601 CCGGACTGACATTCGTGATCAGAACCCCGCGCATGTAAGGCTCCGCGCGAG 6660
Db 6601 CCGGACTGACATTCGTGATCAGAACCCCGCGCATGTAAGGCTCCGCGCGAG 6660
QY 6661 AAGGTCCGTGGAACAATTCCGGGTGACCGGTGCGCGCGGTGAGAAACGGGCTGAG 6720
Db 6661 AAGGTCCGTGGAACAATTCCGGGTGACCGGTGCGCGCGGTGAGAAACGGGCTGAG 6720
QY 6721 GCAACCCGATCATTCGTGTCGTGGAATTCTCTAATCTGTGCGCGCGCATCTTTTGAACG 6780
Db 6721 GCAACCCGATCATTCGTGTCGTGGAATTCTCTAATCTGTGCGCGCGCATCTTTTGAACG 6780
QY 6781 GTGTGTTCCGTGATGACGCGTTCCCGCGCGGTGAGAACTGTGCGGTGAGACCG 6840
Db 6781 GTGTGTTCCGTGATGACGCGTTCCCGCGCGGTGAGAACTGTGCGGTGAGACCG 6840
QY 6841 TTGCGGCGGTGTTTCCCGGTTCGAACTGCGGATTCGTGATTCGCGCGAGTGGAGCG 6900
Db 6841 TTGCGGCGGTGTTTCCCGGTTCGAACTGCGGATTCGTGATTCGCGCGAGTGGAGCG 6900
QY 6901 GGTGCTGACCGGATGATCTGCATCATGAGCGCTCATATGACATCTCTGTAGCATGCT 6960
Db 6901 GGTGCTGACCGGATGATCTGCATCATGAGCGCTCATATGACATCTCTGTAGCATGCT 6960
QY 6961 CCGGCGCGGATTCGACAGGCGCGGAAACGCGCGCATTCGACTGTTTGAACGCTGCA 7020
Db 6961 CCGGCGCGGATTCGACAGGCGCGGAAACGCGCGCATTCGACTGTTTGAACGCTGCA 7020
QY 7021 CATCACGTCGACAGCGCGGATGACACCGACACCGCATGTCGTGTCGCGCATGGAAG 7080
Db 7021 CATCACGTCGACAGCGCGGATGACACCGACACCGCATGTCGTGTCGCGCATGGAAG 7080
QY 7081 GTGCGCGATCAGGGAATGAGCGGTGTCATGACAGACGCAAAAGCTGTCGCGGCT 7140
Db 7081 GTGCGCGATCAGGGAATGAGCGGTGTCATGACAGACGCAAAAGCTGTCGCGGCT 7140
QY 7141 GCGGAAACAGCATGATCTGCTGACGCGTTCAATTGCGCGCGGACCGCTTGAAT 7200
Db 7141 GCGGAAACAGCATGATCTGCTGACGCGTTCAATTGCGCGCGGACCGCTTGAAT 7200
QY 7201 CCGTGCACCGGTGTCGACAGTGAAGCATGCGGACCGCGGTTTGAAGACGCGTGA 7260
Db 7201 CCGTGCACCGGTGTCGACAGTGAAGCATGCGGACCGCGGTTTGAAGACGCGTGA 7260
QY 7261 AAGCGATGACGAGGCTTCTGCTGCGCGCGGACGCGTGAAGTGAAGTGAAGTGA 7320
Db 7261 AAGCGATGACGAGGCTTCTGCTGCGCGCGGACGCGTGAAGTGAAGTGAAGTGA 7320
QY 7321 TCTTCAATCCGGAAGCGTTTATGACAGTTGCGCATGGAAGCGGCTGAGCGAGTTG 7380
Db 7321 TCTTCAATCCGGAAGCGTTTATGACAGTTGCGCATGGAAGCGGCTGAGCGAGTTG 7380
QY 7381 CAGATGCGCGGCTCTCTAATGCGGTTTGGTGGCAACGCGCATCTGTACTGATGCTC 7440
Db 7381 CAGATGCGCGGCTCTCTAATGCGGTTTGGTGGCAACGCGCATCTGTACTGATGCTC 7440
QY 7441 CTGTCCGGAACGAGACGACCGCTGACCTGTGACGAAGGTTGCGGCGCGCGAGCG 7500

Db 7441 CTGTCCGGAACGAGACGACCGCTGACCTGTGACGAAGGTTGCGGCGCGCGAGCG 7500
QY 7501 GACCCGTAACGCACTCGGCGCGGACGTTGATGATGACACCGGACCGCGCGAG 7560
Db 7501 GACCCGTAACGCACTCGGCGCGGACGTTGATGATGACACCGGACCGCGCGAG 7560
QY 7561 GTGTGCGGATCCGCGCTTCAACGAGCGCGCGGACGTTGCGGCGGATGACAGT 7620
Db 7561 GTGTGCGGATCCGCGCTTCAACGAGCGCGCGGACGTTGCGGCGGATGACAGT 7620
QY 7621 GCCCATGCGCGCGGCTCTCTGAGCGCGGCTTCCGGAAGTCTAAGCGCGACGAG 7680
Db 7621 GCCCATGCGCGCGGCTCTCTGAGCGCGGCTTCCGGAAGTCTAAGCGCGACGAG 7680
QY 7681 GAGCGCGGTGCGGATGACAGTGAAGCGCGCATGTCGTCAGACGCGTCCGAGCTGAG 7740
Db 7681 GAGCGCGGTGCGGATGACAGTGAAGCGCGCATGTCGTCAGACGCGTCCGAGCTGAG 7740
QY 7741 ACCGAGCTGAGGTCGCGCTTCCGATCTGATGATGATGATGATGATGATGATGAT 7800
Db 7741 ACCGAGCTGAGGTCGCGCTTCCGATCTGATGATGATGATGATGATGATGATGAT 7800
QY 7801 GCGCTCGGTACCGCGCGCGGACTCAAGGCGGTGACCCGACCGCTCTCGGAC 7860
Db 7801 GCGCTCGGTACCGCGCGCGGACTCAAGGCGGTGACCCGACCGCTCTCGGAC 7860
QY 7861 TCGGCGACCGGATGATGCTGAGACGCGCATGACCGCGCAACGCTCGCGTGAACGAA 7920
Db 7861 TCGGCGACCGGATGATGCTGAGACGCGCATGACCGCGCAACGCTCGCGTGAACGAA 7920
QY 7921 CAGCGCTGACCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGAT 7980
Db 7921 CAGCGCTGACCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGAT 7980
QY 7981 CTGCTGAGGTCGTCGCGGACGTCGCGGCAACAGTGAAGGCAACGCGCTCTGAGCGTC 8040
Db 7981 CTGCTGAGGTCGTCGCGGACGTCGCGGCAACAGTGAAGGCAACGCGCTCTGAGCGTC 8040
QY 8041 ACCGAGCTTCCGAACTGCGCGGACGATCTTCCGACGACCGCGGACCGCGGCTG 8100
Db 8041 ACCGAGCTTCCGAACTGCGCGGACGATCTTCCGACGACCGCGGACCGCGGCTG 8100
QY 8101 GTGACGAGGTCGCGGACGATCCCGGCTCACTGGAAGCGCGCACCGCGCTG 8160
Db 8101 GTGACGAGGTCGCGGACGATCCCGGCTCACTGGAAGCGCGCACCGCGCTG 8160
QY 8161 GACCGCGGTCGCGGAGTCGACGTCGCGACCGGTGACGAGTGTGTCGCGC 8220
Db 8161 GACCGCGGTCGCGGAGTCGACGTCGCGACCGGTGACGAGTGTGTCGCGC 8220
QY 8221 GCGGCGAACGTCGATCCGAGGTCCTTCAACGATCCGACCGGTTTGAAGCGGTC 8280
Db 8221 GCGGCGAACGTCGATCCGAGGTCCTTCAACGATCCGACCGGTTTGAAGCGGTC 8280
QY 8281 GCGGCGAACGTCGATCCGAGGTCCTTCAACGATCCGACCGGTTTGAAGCGGTC 8340
Db 8281 GCGGCGAACGTCGATCCGAGGTCCTTCAACGATCCGACCGGTTTGAAGCGGTC 8340
QY 8341 GTGCGCACCTGACGAGGTCGTCGCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGG 8400
Db 8341 GTGCGCACCTGACGAGGTCGTCGCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGG 8400
QY 8401 GCTTCCGCGCGGTCGATCAGACGCTGCTGACCGCTGCGCGGCTGCTGACCGCTG 8460
Db 8401 GCTTCCGCGCGGTCGATCAGACGCTGCTGACCGCTGCGCGGCTGCTGACCGCTG 8460
QY 8461 CCGGTGAGCTGTAAGGAAAGAAAGTGAAGGCGGCTGTTTCAATGAGCTGTAAG 8520
Db 8461 CCGGTGAGCTGTAAGGAAAGAAAGTGAAGGCGGCTGTTTCAATGAGCTGTAAG 8520
QY 8521 CCACTGTTCCGAGCTGTCGCGCTGCAAGCGCTTCAAGGCGGCGGACGAGGTAAG 8580

Db	8521	CCATCTGTTGGGGGTGGTGGTCCCGCTTCGAGAGGGCTTTCAGGGCCGGCCCGGACACAGAGTAAG	8580
QY	8581	GGTCGTGCGCTTCGCGCGGCTCTTGACCCGACGATCAACCGATGCGGTCTGACCCGCTGGCC	8640
Db	8581	GGTCGTGCGCTTCGCGCGGCTCTTGACCCGACGATCAACCGATGCGGTCTGACCCGCTGGCC	8640
QY	8641	CGTCGGGTACGACCGTGGAACTTGTGGATGTGCAAGCCACGCGGGCCAGGACATCGTGA	8700
Db	8641	CGTCGGGTACGACCGTGGAACTTGTGGATGTGCAAGCCACGCGGGCCAGGACATCGTGA	8700
QY	8701	GTACATGCGGACCCCTCGACTGAGTTCGACCAAGCCACACACCAATGTCTTGGAGCAACT	8760
Db	8701	GTACATGCGGACCCCTCGACTGAGTTCGACCAAGCCACACCAATGTCTTGGAGCAACT	8760
QY	8761	CCTGGGGCATGACAGACCACTTCACCCCGGACCTTCTTGCGCCCTGATGAGCCCGGACTGGCT	8820
Db	8761	CCTGGGGCATGACAGACCACTTCACCCCGGACCTTCTTGCGCCCTGATGAGCCCGGACTGGCT	8820
QY	8821	CATCGACCGGATGTGTGAGTTCTGCGGCTCTGCGGTCGCCGACTGATCGTCTGGAGGCC	8880
Db	8821	CATCGACCGGATGTGTGAGTTCTGCGGCTCTGCGGTCGCCGACTGATCGTCTGGAGGCC	8880
QY	8881	GCTGACCTTTGCGCGGCCCGGATGTGCGGCGCCCGGGTCAACCGGAAACCCCGGACGCGCGGATGCT	8940
Db	8881	GCTGACCTTTGCGCGGCCCGGATGTGCGGCGCCCGGGTCAACCGGAAACCCCGGACGCGCGGATGCT	8940
QY	8941	GTGGGGTCCGGAAGTTCGCGCACCCCGGAGCCCGGACAGACTTCTGTGCACTGTGTCGCCACA	9000
Db	8941	GTGGGGTCCGGAAGTTCGCGCACCCCGGAGCCCGGACAGACTTCTGTGCACTGTGTCGCCACA	9000
QY	9001	GGAGGTGAGCACCCGGAGAGATCC	9024
Db	9001	GGAGGTGAGCACCCGGAGAGATCC	9024

RESULT 2	AXX25774/c	ID	AXX25774	standard; cDNA; 8160 BP.
XX	AXX25774;			
XX				
XX				
DT	08-JUN-1999	(first entry)		
XX				
DE	S. erythraea erythromycin-synthase gene cluster eryAI-eryK.			
XX				
KW	Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;			
KW	secondary metabolite; eryBI; eryCII; eryCII; hybridisation; probe;			
KW	glycosylation; macrolactone; oleandomycin; ds.			
XX				
OS	Saccharopolyspora erythraea.			
XX				
FH	Key	Location/Qualifiers		
FH	CDS	242..1210		
FT		/*tag= a		
FT		/label= ORF13		
FT		/gene= "eryBIv"		
FT		/product= "dUDP-keto-L-6-deoxyhexose-4-reductase"		
FT		1210..2457		
CDS				
FT		/*tag= b		
FT		/label= ORF14		
FT		/gene= "eryBV"		
FT		/product= "mycarosyltransferase"		
FT		2510..3223		
CDS				
FT		/*tag= c		
FT		/label= ORF15		
FT		/gene= "eryCVI"		
FT		/product= "dUDP-D-6-deoxyhexose-3-N-methyltransferase"		
FT		3308..4840		
CDS				
FT		/*tag= d		
FT		/label= ORF16		
FT		/gene= "eryBVI"		
FT		/product= "dUDP-4-keto-L-6-deoxyhexose-2,3-dehydratase"		
FT		4837..6042		
CDS				

[illegible]

Db 7487 GGTGAGCCGTGGCGGTACACGACGTGTGAAGCCGTGAGAGAACTGCTCGCCGCGGAC 7428
Qy 2508 GCGGATCGATGCTGCGGCGCGGAGCTGAGCAAGTCTGTGGAAGACCGAATGCAAGTCTCGGT 2567
Db 7427 GGGGCTGACAGCCCTGTTCTTCCAGCAGCAAGTCTCACCACTCCGCGAGGTGTGTG 7368
Qy 2568 CCGGGGTGACCCGCGCGCGAGCTGAGCGGTGCGCCCGCGCAGCCCGGGGAAACCGGCT 2627
Db 7367 ACGGGGTGACCCGCGCGCGAGTGTAGGGGTGCGCCCTCAGCTCCGGGAAGCCGCT 7308
Qy 2628 CCGGATCAGGTAGAGCTTCCAGCAGGTGTGATTTGACCGCCACTGCGAATGGCG 2687
Db 7307 CCGGATCAGGTAGAGCTTCCAGCAGGTGTGATTTGACCGCCACTGCGAATGGCG 7248
Qy 2688 TCGGGGCGATCGTTTCCGCGCGAGTCCGCAACAGTGGGGGTGACACCCCGAGCAGGC 2747
Db 7247 TGGAGCGCATGCTCTCGCGCGAGTGGCAGCAGCTCGCGTCCACACCCCGCGCAGGC 7188
Qy 2748 TCTCAACGCGTAACCGAGGTGCGTCTGACATGCGCCGCGGCTCGCTCGGCGCGGTGCA 2807
Db 7187 TCTCAAGGCGGTAGCCGAGGTGATGTGATGCTCCCGGGGTCCGCTCGGCGCGGTGATGCA 7128
Qy 2808 CGAACCGGAGGAGCCCTCCGCAACTCGGTCCGTTCCGCTCGACCACTGCGCTGT 2867
Db 7127 CGAACCGCACAGGCCCTCGCGCAGCTGTTGCGCTCGAGTCCGACAGCGCGCTGT 7068
Qy 2868 CCGGACCGCTGTAGTCTTCCGCGAGCTGATGCAAGTGTGAGCGCGCGGTGTGTGACT 2927
Db 7067 CCGGCGCGCTGTAGTCTTCCGCGAGCTGATGCAAGTGTGAGCGCGGTGTGTGACT 7008
Qy 2928 CGTTGAGTTGCGGATGAGTGTGACCAAGTGTGAGTGTGAGCGCGCGCGCGCGAGCA 2987
Db 7007 CGTTGAGTCTCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGCGCGCGCGCGCGAGCA 6948
Qy 2988 TGAATGTTGAGAGCGCGAGCGAGCGAGTCCGCGCTTCCGCGCGAGTGTGAGAGT 3047
Db 6947 TGAATGTTGAGAGCGCGAGCGAGTCCGCGCTTCCGCGCGAGTGTGAGAGT 6888
Qy 3048 CCGGAGGTTCTTCTGACCCGTTCCAGGCGCGAGCTTCCGCGGTGTGTGAGTGTGACT 3107
Db 6887 CCGGAGGTTCTTCTGACCCGTTCCAGGCGCGAGCTTCCGCGGTGTGTGAGTGTGACT 6828
Qy 3108 CGTGTGTTGACCGCGTACAGGAGGTGTGACCGCGCGAGTGTGACAGGCGCGCT 3167
Db 6827 CGTGTGTTGACCGCGTACAGGAGGTGTGAGTGTGAGTGTGAGCGCGCGCT 6768
Qy 3168 GCGCGCGAGCTGTCTGTGAGTGTGAGGCGAGGCGTGTGTGAGTGTGAGCGTGTGAGT 3227
Db 6767 GCGCGTGTGAGCTGTCTGTGAGTGTGAGGCGAGGCGTGTGTGAGTGTGAGTGTGAGT 6708
Qy 3228 CCGCGCGGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3287
Db 6707 CCGCGCGGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6648
Qy 3288 CCGCGAGAGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3347
Db 6647 CCGCGAGAGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6588
Qy 3348 GATGTGTGTGCGGCGACCACTGCGGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3407
Db 6587 GATGTGTGTGCGGCGACCACTGCGGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6528
Qy 3408 AGAATGTGACAGCGGAATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3467
Db 6527 AGAATGTGACAGCGGAATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6468
Qy 3468 CCGGCGTCCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3527
Db 6467 CAGGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6408
Qy 3528 TGTCTCAAGTGTGCGCGCGCGCGCTGTGCTCAACCGCGGTGTGAGTGTGAGTGTGAGTGTGAGT 3587

Db 6407 TGAACCAAGTACTTGTCCCGCGGACCGTGTGTGACCGCGGTGTGAGTGTGAGTGTGAGT 6348
Qy 3588 CGAACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3647
Db 6347 CGAACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6288
Qy 3648 GCTCAACAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3707
Db 6287 GCTCAACAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6228
Qy 3708 CTTGTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3767
Db 6227 CTTGTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6168
Qy 3768 CCGGACCAACGCGCGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3827
Db 6167 CCGGACCAACGCGCGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6108
Qy 3828 CCGCGGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3887
Db 6107 GCGCGGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6048
Qy 3888 GAGATCAAGAACCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3947
Db 6047 GAGTGTGAGAACCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5988
Qy 3948 GTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4007
Db 5987 GTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5928
Qy 4008 CCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4067
Db 5927 CCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5868
Qy 4068 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4127
Db 5867 GGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5808
Qy 4128 ACTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4187
Db 5807 ACTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5748
Qy 4188 TCCACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4247
Db 5747 TCCACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5688
Qy 4248 GTGACGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4307
Db 5687 GTGACGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5628
Qy 4308 TCCGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4367
Db 5627 TCCGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5568
Qy 4368 ATCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4427
Db 5567 ATCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5508
Qy 4428 ATGCGCGGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4487
Db 5507 ATGCGCGGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5448
Qy 4488 GCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4547
Db 5447 GCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5388
Qy 4548 GAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4607
Db 5387 GAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5328
Qy 4608 TGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4667
Db 5327 TGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5268

QY	4668	CCGACGATCGCGCCGGTCTGCGGGGTGA	CAGCGCGCGACGTGGTCCGGTGCACAGA	4727	
Db	5267	CCGATGATCGCGCCCGTCCGGGTGTCA	CCAGCGACCGACGTGCTCGGGGTGAGACAGG	5208	
QY	4728	CCGGTCCGCGGGGTGACGTCCGACAA	CAACCGGTGTGAAGTCCGAACCGACTCCCGCGTGC	4787	
Db	5207	CCGGTCTCGGGGTCCACTCGCAAA	CAACCGGTTCAGGCCCAACGCTCGCGCGTGC	5148	
QY	4788	GGCGTGGCCGGAAGGTCAATTGA	CGGCATGATCTTCAACCGGTGAGTCAACCGGC	4847	
Db	5147	GGCGTGGCCGGAAGTCAATGAA	AGCATTAACCTCGCGACAGTCCGCTCGCGCGC	5088	
QY	4848	AGCAGCAATTCCAGGCGACGCGTGG	CGTTGCAAGTGGCGATGACAGTGCCTTACC	4907	
Db	5087	AGCAGCAATTCCAGGCGACGCGTGG	CGTTGCAAGTGGCGACGAGTGGACACCGCGC	5028	
QY	4908	AGGTCCGCGACCCCGGACCTCGAA	CTCCCGCACAGGGGCCCGCGGTGTGAAGCAATTGG	4967	
Db	5027	AGGTCCGCGACCCCGGACCTCGAA	CTCCCGCACAGTGGTCCGCGTGTGTGACCAATTGG	4968	
QY	4968	TTGTTCAGGGCCCACTCCAGGCG	GGGCCAGAAACCGTGC	CGGTGACGATCGTGGCGCG	5027
Db	4967	TTGTTCAGGCGCCACTCCAGGCG	GGGCCAGAAACCGCTCCCGGTCCCGACG	CGGTGGGCTTG	4908
QY	5028	CCCAAGTGCAGGGGGTGCAGAA	CGCCTCGGGAACCGCGAAAGATCGCCAGATCGGT	CGGT	5087
Db	4907	CCCAAGTGAAGGGTGTGCAGAA	TTGCTCGGGGCGCGGAAGATCGCACGGT	CGGTGAGC	4848
QY	5088	ACGCGCTTCA	CGCGGTCCGCGTGTGTGAACCGCGCAACCGCAGCGCAACAGGCTTGCGCA	5147	
Db	4847	GGCGCTTTCATCCCGCGGTCTCGGTGTGA	CGGCGCGACCGACCGCATGATGTCTGCGCA	4788	
QY	5148	GCTGGAATGTTGACGTAAGTGG	CGTGGCGGAGACAGTTGCGTATCTGGCCGAAGGTCAATCC	5207	
Db	4787	GCTGGAATGTTGACGTAAGTGG	CTGTGCCAGAGAGTAAAGTATCTGTTCGAAGGTCAATCC	4728	
QY	5208	ACCGGTGATCGGGGCGACCGCGCT	CGTGAAGTCCGCGGGACCTCGACAGCATGTAACC	5267	
Db	4727	ACCGGTGATCGGGGCGACCGCGCT	CGTGAAGTCCGCGGGACCTCGATCAAGTGAATACC	4668	
QY	5268	GATTCTGTTGGCGTGAACA	CCGCCCACTCTCGAAGTCCAGGCGTGTGAAGCGATCT	5327	
Db	4667	GATTCTGTTGGCGTGAACA	CCGCCCACTCTCGAAGTCCAGTGTGAGCGAAGCT	4608	
QY	5328	GGGTACGGGGACCGTCCAGCA	CGTGAAGTGGAGGAGTGGGCGGATCTCCCGGGACGT	5387	
Db	4607	GGAGACGGGACCGTTCAACA	CAAGTACATTAAGCCGGTGGAACTCTCCGGCGGT	4548	
QY	5388	CGGTGATGTTGTCGGGCTGA	CAGTGCACCGTCCGCAACTCGGCGACGTCGTGGCAC	5447	
Db	4547	CGGCTTAAGTTGTCCGGGCT	GGCAGTGCACCGTCCGCGACGTCATCCCGC	4488	
QY	5448	CCGCTTCGGTGCAGGTGCAC	AGACCGCGTGAAGGTGCCGTGCATCTCTTTGACAGGA	5507	
Db	4487	CCGCTTCGGTGCAGGTGCAC	AGACCGCGTGAAGTGCATCCCGCATGTCTTTGACAGGA	4428	
QY	5508	GGGCGACGACATGCGGTGTT	GGGGGGGAGAGACGCGTTGATCCAGGACCTCAACCTCCG	5567	
Db	4427	GGGCGACGAGCGCGTTGTT	GGGGGGGAGAGACGCGCTGCATCCAGGATTCGACCTCGC	4368	
QY	5568	GGTGACTGGTCTGCACGAG	AGACCGCGGAAGATCTGAATTACTTGTCTCGTGTGGA	5627	
Db	4367	GGTTCGCTGGTCCCGAC	CGGTGAACCGCGGAAGACTGAATTACTTCTCTCTGTCGCA	4308	
QY	5628	TGCGGTCTGTCGCGCGGAT	TCACCCGTCGGTGCATCGGCAACGCGCGGGTGCCTT	5687	
Db	4307	TGCGGTCTGTCGCGCGGAT	TCACCCGTCGGTGCATCGGCGAGCGGGCGCGCGCT	4248	
QY	5688	GGAACCACTCA	CGTGTGGTGCATGTCAGATTCACACTCAGCAGGATTCATTTGT	5747	
Db	4247	GCAACCACTCGCGCT	CGCGCGACAGTGCAGATTCACACTCAGCAGGCTGTTCATGTTGT	4188	

QY	5748	GGACCCGGGGTCCCGCCGCCACGAAAGGAAGCAGAGAGTGGGGTCTCGAAGAAAGCCCTGG	5807
Db	4187	TGGCCCGGGGGTCCCGCCCGATGAACAATGGACAGACCCGGGCTTGAAAGAGCCCTTCGG	4128
QY	5808	GCAGTCCGCTCCAGATACCAGCCGAGCCGAGCCCGGCTTCACCTTCGATGGATGACAGCGCA	5867
Db	4127	GCAGGCGCGAGGACAGTCCGTCCGGCCCGGCTCCGCTCCAGCGCGGTCCGGACCGAGGCCA	4068
QY	5868	GCACGAGACCGAGGTCCATGTTCCACACGTTGTCGATGAGCGAGATCGCCCGACGCTGGG	5927
Db	4067	GCACGGTACCGAGGTCCATGTTCCACACGTTGTCGATGAGCATGCGCCGACGCTGGC	4008
QY	5928	CGAGGGTGAAGCACCTCTGCAGTTCCGGGGTCTCCGGCCGGGCTCTGGAACCTCCAGCACCA	5987
Db	4007	CGAGGGTGAAGCACCTCTGCAGTTCCGGGGTCTCCGGCGAGGTCTGGAACCTCCAGCACCA	3948
QY	5988	TGTTGCGGGTTCGTTTGCGCAGAAACCGAAGAACCTGCTCGACTGAGAGACGTTCAGCA	6047
Db	3947	TGTTCCGGTTTGCGCTTGCGCAGAAACACCGCCCTCTCGACTGAGAGACGTTCAGCA	3888
QY	6048	GAATCCGGTGGAGGCGGGTCCCGTCGAAGTACTCGATGAACTGAGCGCGGATCCGTTGT	6107
Db	3887	GAATCCGGCTCGGGCGGGTCCCGTCGTTGAAGTACTCGATGAACCGGACTTCGAGCGCGGT	3828
QY	6108	GGAACCTCATATGATGTTGTGTGGGGTGGCTCTGACAGGGTCCGCGAGAGCTGACACCGCTTGA	6167
Db	3827	GGAACCGCTGATGATGTTGTGTGGGGTGGCTCTGACAGGGTCCGAGAGAGCTGACACCGCTTGA	3768
QY	6168	TGTTGCGCAGGTTCCGGGCTTGAGCTGCAACAGAGGGTCAAGTACGCGCTGCACCTCCGCA	6227
Db	3767	TGTTGCCGGCTCCGGCTTGAGCTCTGAGCGCGCAGACAGTGCAGCACACCTGCACCTCTTGA	3708
QY	6228	CGATCAGACCGAGGAACCCGATCTCGGGTTGGACGATGATGGATTGGATCCAGTCCGCTC	6287
Db	3707	CGATGAGGCGCAGAGGAAGCGGATCTCGGGCTGCAAGATGAGGCTGGATCCAGTCCGCGC	3648
QY	6288	GCCATCCGAAGTTGGTCCGACGTTGCAAGGCCCTCTGATGAGAAAGAACGCCCGGAGTCTG	6347
Db	3647	GCCAGCCCGAAGTTGGTCCGACAGTGCAGAGGCCCTCTGATGAGAAAGAACGCCCGGAGTCTG	3588
QY	6348	GCGCGAGCGGACCGTCCCTCCGGGGTGAACAACGACGCGTCCATGAGTGTCTGAAAGGACCTC	6407
Db	3587	GCGCGAGGCTGCGGCTCTCGGGCTGGAACGACGCGGTCCATGAGCGCTGAACGGACGCG	3528
QY	6408	GGTGCACCTCGAGCGGATGCTCGGCGGCGCGCGGAGGCCAACAGTCTGTGACGTCGTCCG	6467
Db	3527	GCGCGACCTCGAGCGCGGCTCTCGGCGCGGCTGTGGGGTCAAGCAATCGGGGTTTCCGCG	3468
QY	6468	TGGCGGTGGAGGTCCGCGCTGGCCGAGTCAAGAAACGTATTGCGGATTTGTGGATTCCG	6527
Db	3467	TGGCGGTGGAGATTCCGCGCTGGCCGAGCAGAAACGGTTGCGGCTGTGCGAGGATTCCG	3408
QY	6528	CAGTGGCATGACCGTTGACCCGATCCCCCATACGCTCTCCCGGATGATGCTGGGCGGCTC	6587
Db	3407	GAATGTCCTGGTCCCGGTCGATCACCATTCGACTCTCTTGCGGTGTCTGACAGGCT	3348
QY	6588	CGTGGGTACCGCCCGGACTGACATTCGTCGATCAAGACC-CGCCAGTGTAGGGCTCC	6646
Db	3347	CGGTTGCTGCGCGCGGGC-----GTTGTGATCAAGACCGGCACTCATGTAAAGGTC	3293
QY	6647	GCCCGCGACCGGAGAAAGGTCC	6667
Db	3292	GGTCAACGCGGAGAAAGGCC	3272
RESULT 3			
AA772685/c			
ID AA772685 standard; DNA; 8051 BP.			
XX AA772685;			
DT 18-Sep-1997 (first entry)			
XX			

RESULT 3
AAT72685/c
ID AAT72685 standard; DNA; 8051 BP.

DT 18-SEP-1997 (first entry)

OY	3468	CGGGCTCCCGCGGTAACGCAAGCTTCTCAACGAGCCGAGGGGCTTCAAGGAGCTGATGTGT	3527
Db	6304	CAGGCTTGGCGGTAGACCGCGCGCTTCAGTGGGCTTGGGGGATCCAGGAGGCTTATGCTGT	6245
OY	3528	TGCTCTCAGTATCTTGACCGCGCGGAGGCTCGATCTCAACCGCGGTGACCAGCTGCGGAGATGCAT	3587
Db	6244	TGGAACAGTACTTGTCCCGCGCGGACCGTGTCTCAACCGCGGTGTCCCAAGGTCCGGGATGGCG	6185
OY	3588	CGAAACAGGTGAGGAGCTCCGCGAACCCCGAGCCGGTCCAGGTGGAACCGACGGCGCATCT	3647
Db	6184	CGAAACAGGTGAGGAGGAGCGGAATCCCGGCGGCTCAAGGCGCAGGCGCGACGCGTCT	6125
OY	3648	GCTCTCAACGGGAGTGAACGGCGGAGCGCGCTGATGTGGGGGTGTGAGTTCCGCGGACGCGGACCG	3707
Db	6124	GCTCCAGGGGGGTGAACGGGTGCGCCCTGTATGTACGGCGGAGGTGTGGCGACGCGGGCGGG	6065
OY	3708	CCTGTGCTTCCGCGGAGTGTCCGATCGGCGACCGGATGAGGAGGAGGTGATCCTCGAGCTGACCG	3767
Db	6064	CTTGACAGTGGAGCGTCAACCTTCAACGGCGCGCGGTGACCGCGGAGACCTTCGAGCTACGCG	6005
OY	3768	CCCGGACCAACGGCGCGGCGGATCGGCGGTGCGGTGCGGTCCGCGATCTTCGTGGTGA	3827
Db	6004	CCCGGACCAACGGCGCGGCGGATCGAAGTCCGAGCGCGCGCGGTCCGCGACCCCGCTTCT	5945
OY	3828	CGGCGGTTCGGCTGAGGAATGTCTTTCATTCACGTGTCCAAACCTTCTGCGGTCTGTGGTGC	3887
Db	5944	GGGCGGTGTCCGTACGAGTGTGTTCATCCGGGAGCGAACCTCTTGACACTTGTGTGGGCA	5885
OY	3888	GAGGATTCACGAACGGTTCGTTTCGCGTGTCCACTCCGCGGTGATCAACGACCGCTG	3947
Db	5884	GGTGCTCACGAACGTTTGGCGGTCTCTGTGTCCATTTGGCGGTGATCATGTCGCGCTGC	5825
OY	3948	GTGTGTGCTAGTCGGATGACGTTCGCAACCCGGCGGATGTCTCATTTGACACCGGAGGG	4007
Db	5824	GTGGCGGCGAGCGGATGATGTTCGACACCCGCGGATGTCTCGCTTGACACCGCGGGG	5765
OY	4008	CCGGTCGGGAGGCGGACGCCCTTCCGCGGAGCCACTCCGCTCTGTGTACAGCGCGGAGGGT	4067
Db	5764	CCGGTGGGCGAGCGGACGCCCGGTCCGAGAGCTGTGTGGTGTCTCCAGCGCGGAGGGC	5705
OY	4068	GAGTCCGTGCGGTAGGAGCGTCACTGTGTGAGGAGGCGGGGAGAGTAAAGGCTGTGCGAGC	4127
Db	5704	GAGTCCGTGCGGTAGGAGCGTCACTGTGTGAGGAGGCGGGGAGAGTAAAGGTTGTGCGACA	5645
OY	4128	ACCTTCTCCGCGCGGAGATGCGCTGCAAGCCGGTCAAGGTGATGCGGTGGCGGTGCCG	4187
Db	5644	ACCTTCTCCGCTCGGAGATGCGCTGCAACTGTGTGCGGTGATGCGGTGGCGCGGAG	5585
OY	4188	TTCACCAAGATGATCAACGTATCTGTGATGTTGCTCTCTGTGGGGCGGGAGCGAATGACG	4247
Db	5584	TTCACCGAAGATATACGTACTGTGATGTTGTCTGTCCAGGATTCGAAGCGGTGACG	5525
OY	4248	GTGACGCGCGGTAAGTCCGCGAGCTCGCTGTGTGTGAGAGCGGAGTTGACCCGGTGTGC	4307
Db	5524	GATATGCGCGCGACCTCCCGAGAGCTCGTCCGAGTGAAGCGGTGTGTAAGCGGTGTGTGC	5465
OY	4308	TCCCTGTGCTCGGCGAACGCGTTCAGGAGGAGTTGAAGCCCATGCGCGGCGCACTTCGCTC	4367
Db	5464	ACCTTGCTTCGCGGAAAGGCGTTCAGGAGGAGTTGAAGCCCATTCGCGCGCGCACTCGCTC	5405
OY	4368	ATCTTGGCGGTGATCCCATCTCCGGTGAACACTTGTTCGGGCGCGATGCGCAAGTTGTGC	4427
Db	5404	ATCTTGGCGGTGATCCCATCTCCGGTGAACAAGTGTTCGGGTGCATCCCAAGTTGTGC	5345
OY	4428	ATTGCGCGGATCCGTTTCGCGACGAGCGCGTGTGCGTGAACAACGCGCCCGCTTCGAAG	4487
Db	5344	ATTGCGCGGATCCGTTTCGCGACGAGCCCGTGTGCGTGAAGATGGCCGCCCTTCGAAC	5285
OY	4488	GGCGTGAACGCTTGTGATGCGTGAAGCTGAACACTTCGCGGTCAACGGAATCCGCCAC	4547
Db	5284	GAGGTGAACGCTTGTGATGCGTGAAGCTGAACACTTCGCGGTTCGCAAGGCGCGAC	5225

QY	4548	GGAGCTGCACCCGTCGTGCAAGCCAGGAGCTGGAGCGCGCTGCAGAAAGAAAGCTGACCTGG	4607
Db	5224	GGCGCCCGCGCGGGTGCAGCCCAAGCGCTGGCGCGCTGCAGAAAGAAATTGGACTTGG	5165
QY	4608	TGGTCGCGGGGAACTCTTCGCAAGCGCTTCCACAGATGCTGATGGAGCCCAAGATGTACG	4667
Db	5164	TGCTTCGGCGAGATCTTTCTTCAGAGGCTTCGACCGGAGCGGGCTTGCCCAAGATGTACG	5105
QY	4668	CCGACGATTCGGCGCGGCTCTGCGGGGTGACGAGCGGGGACGCTGTCTCGGATGCACGA	4727
Db	5104	CCGATGATTCGGCGCGGCTCTCGGGGTGTACACGACGACGCTGTCTGGGATGCAGACGG	5045
QY	4728	CCGGTCGCGGGGTGCAGATGTGCAGAAACACGGGTGTGATGTCCAGGCCAGACTTCGCGGTGC	4787
Db	5044	CCGGTCTTCGGGGGTGCAGTGTGCAGAAACACGGGTTCAGGCCCAAGCCAGCTTCGCGGTGC	4985
QY	4788	GCGGTGGCCGCGAAGGTCAATTGACGGGATGATCACTTTCACCGGATGAGTACACGACCGCG	4847
Db	4984	GCGGTGGCCGCGAAGCTTCATGAAAGGATGACACTTCGCGGACAGCTGTGCTGCAGCGGC	4925
QY	4848	AGCACCAATTCAGAGGACGAGTGGCGGTTCGAGAGGTGGCGATGCAGTGCCTGACCCGACC	4907
Db	4924	AGCACCAATTCAGAGCGGACCGGTGCGGTTCGAGAGGTGGCGAGCTGTGGCGGACACCGCC	4865
QY	4908	AGGTTCGCGGACCCCGGCGCTTCGAACTCCCGCACAGGGGCGCGCGCTTGTGTGACCAATTG	4967
Db	4864	AGGTTCGCGGACCCCGGCGCTTCGAACTTCGCGCACAGTGTTCGCGGTGTGTGACCAATTG	4805
QY	4968	TTGTTCAGAGGCGCACTCCAGCGCGGGCCAGGAAACCGCTGCGCGGTGCAGATCTGTCCGCGG	5027
Db	4804	TTGTTCAGAGGCGCACTCCAGCGCGGGCCAGGAAACCGCTTCGCGTTCGCGACGATTCGCTG	4745
QY	5028	CCCAAGTCAGAGGGGTGCAGGAAACGCTTCGAGACCGCGAAGATTCGACAGATCGATCGAT	5087
Db	4744	CCCAAGTCAGAGGGGTGTGAGGAATGCTTCGAGGCGCGCGAAGATTCGACAGATCGATCGAC	4685
QY	5088	ACGCGCTTCACAGCGCTGCCTCCGTTGTGATGACGCGCGACGCGCAGGCGACAGAGTTCGCGCA	5147
Db	4684	GCGCGCTTCATCCGCGCGGTCTCGTGTGTGACGCGCGAGCGCGCACGCGCATATGTCTGCGCA	4625
QY	5148	GCTGGAATTTGACAGTATGATGCTGTGGCGGAGCAATTCGATGATCTGGCCGAGAGGTCAATCC	5207
Db	4624	GCTGATTTTGAACGTATGATGCTGTGGCGGAGCAATTCGATGATCTGGATCGAAGGTCAATCC	4565
QY	5208	ACCGGTGATCGAGGGGACACCGCGCTGTGTGAAAGTCCGCGGGGACCTTCGACGAGCATGTATCC	5267
Db	4564	ACCGGTGATCGAGAGCGGACACTGGCGTTCGAAAGTCCGCGGGGACCTTCGATCAGCATGTATCC	4505
QY	5268	GGTTCTCGTTCGAGTGAACCGGCCACCTCTTCGAGATGTCAGAGCGCTCGTATCGGATCT	5327
Db	4504	GGTTCTCGTTCGAGTGAACCGGCCACCTCTTCGAGATGTCAGAGCGCTCGTATCGAGCGCT	4445
QY	5328	GGGTACGCGGGAACCTTCAGACGATATGTCGAATGATTCGGCGGGTGTTCCTCCGAGGACGT	5387
Db	4444	GGAGACCGCGGACCTTTCACACGTAATGTCACATATGGCCGAGTCGAACTCTCCGAGCGGT	4385
QY	5388	CGGTGATGTTTGCAGCTGCAGATGACACCTTCGCGCCCACTTCGAGCAGAGCTGTGGCAC	5447
Db	4384	CGGCTATGTTTGTCCGGCTTGGCAGATGACACCTTACGGCGCCAGCTTCGAGCAGAGCTGTGC	4325
QY	5448	CCGCTTCGATGCGAGATGACAGACCGGTGAGAGGGTGCAGTCTCTTTCGACAGAA	5507
Db	4324	CCGCTTCGATGCGAGCTGACAGACCGGTGACAGAGCTTCGCGCATGCTTTCGACAGAA	4265
QY	5508	GGGCGAGCAATCCGTGTGTTGGCGGGGGAGAGCAGCGTTCATTCAGGACCTTCACCTTCGC	5567
Db	4264	GGGCGAGCAGGCGCTTGTGTGGCGGGCGAGACAGCGCTGATTCACAGATTTGACCTTCGC	4205
QY	5568	GGTGAATGTGTGTCACAGAGACCGCGAAGATGTGAAATTAATTCTTGTGCTCGATTCGAA	5627
Db	4204	GGTGTGTGTGTCCACCGGTGACCGCGAAGAGCTGCAAATTAATTCTTCTCTCGATTCGAA	4145
QY	5628	TGCCCTGTTCGCGCGGATTCACCCGCTGTGCGATGTTCGAGCACAAGCGCGAGTTCGCTT	5687


```
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 1"
FT /note= "polyketide synthase; encodes AAB8212"
FT 12505. .13470
FT /*tag= m
FT /gene= "megaA"
FT /function= "AT-L"
FT 13576. .13791
FT /*tag= n
FT /gene= "megaA"
FT /function= "ACP-L"
FT 13849. .15126
FT /*tag= o
FT /gene= "megaA"
FT /function= "KS1"
FT 15427. .16476
FT /*tag= p
FT /gene= "megaA"
FT /function= "AT1"
FT 17155. .17694
FT /*tag= q
FT /gene= "megaA"
FT /function= "KR1"
FT 17947. .18207
FT /*tag= r
FT /gene= "megaA"
FT /function= "ACP1"
FT 18268. .19548
FT /*tag= s
FT /gene= "megaA"
FT /function= "KS2"
FT 19876. .20910
FT /*tag= t
FT /gene= "megaA"
FT /function= "AT2"
FT 21517. .22053
FT /*tag= u
FT /gene= "megaA"
FT /function= "KR2"
FT 22318. .22575
FT /*tag= v
FT /gene= "megaA"
FT /function= "ACP2"
FT 22867. .33555
FT /*tag= w
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT /note= "polyketide synthase; encodes AAB8213"
FT 22957. .22957
FT /*tag= x
FT /gene= "megaA"
FT /function= "KS3"
FT 22957. .24237
FT /*tag= y
FT /gene= "megaA"
FT /function= "KS5"
FT 24544. .25581
FT /*tag= z
FT /gene= "megaA"
FT /function= "AT3"
FT 24544. .25581
FT /*tag= aj
FT /gene= "megaA"
FT /function= "AT5"
FT 26230. .26733
FT /*tag= z
FT /gene= "megaA"
FT /function= "KR3 (inactive)"
FT 26230. .26733
FT /*tag= ak
FT /gene= "megaA"
FT /function= "KR5"
FT 26998. .27258
FT misc_feature
```

```
FT /*tag= aa
FT /gene= "megaA"
FT /function= "ACP3"
FT 26998. .27258
FT /*tag= al
FT /gene= "megaA"
FT /function= "ACP5"
FT 27393. .28590
FT /*tag= ab
FT /gene= "megaA"
FT /function= "KS4"
FT 27393. .28590
FT /*tag= am
FT /gene= "megaA"
FT /function= "KS6"
FT 28897. .29931
FT /*tag= ac
FT /gene= "megaA"
FT /function= "AT4"
FT 28897. .29931
FT /*tag= ad
FT /gene= "megaA"
FT /function= "DH4"
FT 29953. .30477
FT /*tag= ae
FT /gene= "megaA"
FT /function= "ACP6"
FT 32257. .32799
FT /*tag= af
FT /gene= "megaA"
FT /function= "RR4"
FT 31396. .32244
FT /*tag= ap
FT /gene= "megaA"
FT /function= "ACP6"
FT 32257. .32799
FT /*tag= ag
FT /gene= "megaA"
FT /function= "ACP4"
FT 33052. .33312
FT /*tag= ah
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 3"
FT
```

Query Match 29.2%; Score 2606.8; DB 4; Length 47981;

Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;

Matches 2608; Conservative 0;

```
QY 6415 CTCGAGCCGATGCTGGCGCGCGCGGTGGCCAAACGATCGTGAAGTCTGCTGGTGGCGGT 6474
DB 1 CTCGAGCCGATGCTGGCGCGCGCGGTGGCCAAACGATCGTGAAGTCTGCTGGTGGCGGT 60
QY 6475 GGGAGGTCCGCGCGCGCGCGATGAGGAAACGATTCGCGATTGCGATTCCGAGTCCG 6534
DB 61 GGGAGGTCCGCGCGCGCGCGATGAGGAAACGATTCGCGATTGCGATTCCGAGTCCG 120
QY 6535 ATGACCGTTGACCCGATCCCGCATACGCTCTCCCGTGAATGTCGTGGCGGTCCTGTCGG 6594
DB 121 ATGACCGTTGACCCGATCCCGCATACGCTCTCCCGTGAATGTCGTGGCGGTCCTGTCGG 180
QY 6595 TACCGCCCGACCTGACATTCGTCGATCAAGACCCCGCCAGTGTAGGAGCTCCGCCCGCA 6654
```


Db 181 TACCGCCGACATGACATCTGCGATCAAGACCCCGCCAGTGAAGGCTCCGCGCCGCA 240
Qy 6655 CGGAGAGAGGTCGGTGCAGCAACTTCCGGGTGACCCGTCGCCCGCGTGGTGAACGGGC 6714
Db 241 CGGAGAGAGGTCGGTGCAGCAACTTCCGGGTGACCCGTCGCCCGCGTGGTGAACGGGC 300
Qy 6715 GTGAGAGCAACCCGATCATCTGCTGCGTGAACCTTCTTAATGTCGCGCGGCACTTTTC 6774
Db 301 GTGAGAGCAACCCGATCATCTGCTGCGTGAACCTTCTTAATGTCGCGCGGCACTTTTC 360
Qy 6775 TGAACCGGTGTTCCGTGTATGACGCGTTCCCGCCCGTCTGGAACCTGTGCGGACT 6834
Db 361 TGAACCGGTGTTCCGTGTATGACGCGTTCCCGCCCGTCTGGAACCTGTGCGGACT 420
Qy 6835 GACCGGTGCGGGGTGTTTTCCGCCCTTTCGGAACCTGCGGATTCGTGATTCGCGAGGTG 6894
Db 421 GACCGGTGCGGGGTGTTTTCCGCCCTTTCGGAACCTGCGGATTCGTGATTCGCGAGGTG 480
Qy 6895 GAGACGGGTGCTGACCGGGATGATCTGAATGAGCGCTCAATGACATCTTTAG 6954
Db 481 GAGACGGGTGCTGACCGGGATGATCTGAATGAGCGCTCAATGACATCTTTAG 540
Qy 6955 CATGTCGCGCGCGAGGTCGACAGGCGCGAAACGCGCGCATCCAGCTGTTGACGA 7014
Db 541 CATGTCGCGCGCGAGGTCGACAGGCGCGAAACGCGCGCATCCAGCTGTTGACGA 600
Qy 7015 CGTGCACATCAACGCTGACAGCGCGCATGACACGACCGCATCTGTGTCGCGACT 7074
Db 601 CGTGCACATCAACGCTGACAGCGCGCATGACACGACCGCATCTGTGTCGCGACT 660
Qy 7075 GGAAGGTTGGCGGATCAAGGAAATGCGCGTCTCATGACAGACGCGCAACGCTGTC 7134
Db 661 GGAAGGTTGGCGGATCAAGGAAATGCGCGTCTCATGACAGACGCGCAACGCTGTC 720
Qy 7135 GGGCTGCGGAAACAGATCATCTGTCAGCGCTTCAATGTCGCCGCGCGCAACGCTT 7194
Db 721 GGGCTGCGGAAACAGATCATCTGTCAGCGCTTCAATGTCGCCGCGCGCAACGCTT 780
Qy 7195 GGAATTCCTGTCACCGGTGTCCGACGTGACGATGCGGACCGGGTTTCAGACAGCA 7254
Db 781 GGAATTCCTGTCACCGGTGTCCGACGTGACGATGCGGACCGGGTTTCAGACAGCA 840
Qy 7255 GGTAGTAGGCGATGACGCGTTCTGTCTGCGCGGACGCGTGGCACTAGGTGAATCCG 7314
Db 841 GGTAGTAGGCGATGACGCGTTCTGTCTGCGCGGACGCGTGGCACTAGGTGAATCCG 900
Qy 7315 TCAAGTCTTCAATCCGAGACGTTCTATGTCAGTTGTCGATCGAAGCGCGTGGCCG 7374
Db 901 TCAAGTCTTCAATCCGAGACGTTCTATGTCAGTTGTCGATCGAAGCGCGTGGCCG 960
Qy 7375 GAGTTGCAGATGCGCCGCGGTCTTACTGAGGGGTTCCGTGCAACGCGCATCTGTACTCG 7434
Db 961 GAGTTGCAGATGCGCCGCGGTCTTACTGAGGGGTTCCGTGCAACGCGCATCTGTACTCG 1020
Qy 7435 ATGCTCTGTTCCGGAAGGACGACGACCCCTTGACTGTGTAAGAGGTTGCGGCGCCG 7494
Db 1021 ATGCTCTGTTCCGGAAGGACGACGACCCCTTGACTGTGTAAGAGGTTGCGGCGCCG 1080
Qy 7495 GGAACGGGACCGTACGACGTCGAGCGCGAAGCTGAGGTGTCGTTGACACCGGACGCG 7554
Db 1081 GGAACGGGACCGTACGACGTCGAGCGCGAAGCTGAGGTGTCGTTGACACCGGACGCG 1140
Qy 7555 GCCGAGGTGTCGCGCATCCGAGCTTCAACCGACGCGCCGCGACGCTGCGCGGTGATG 7614
Db 1141 GCCGAGGTGTCGCGCATCCGAGCTTCAACCGACGCGCCGCGACGCTGCGCGGTGATG 1200
Qy 7615 CAGGTGGCCACTGCGCGCGCTCTGAGCGCGCCCTTCCGGGAGTTCTACGCGCGC 7674
Db 1201 CAGGTGGCCACTGCGCGCGCTCTGAGCGCGCCCTTCCGGGAGTTCTACGCGCGC 1260
Qy 7675 ACCGAGACGCGCGTCTGATGACGTGACGCGCATGTCGCTCCAGACGCGGTGCGCAGG 7734

Db 1261 ACCGAGACGCGCGTCTGATGACGTGACGCGCATGTCGCTCCAGACGCGGTGCGCAGG 1320
Qy 7735 CTGTTGATCCGAGGTCGGGTGCGCGCTTCCGATTCCTGTGAACGACTTTCGCCGGAGGTCCG 7794
Db 1321 CTGTTGATCCGAGGTCGGGTGCGCGCTTCCGATTCCTGTGAACGACTTTCGCCGGAGGTCCG 1380
Qy 7795 GTGCTGGCGCTCGGTACCGCGCCCGCATCAAGGGGCTGGAACCCCGACGCTCCGGTCC 7854
Db 1381 GTGCTGGCGCTCGGTACCGCGCCCGCATCAAGGGGCTGGAACCCCGACGCTTCGGTCC 1440
Qy 7855 TGAACCTGCGGACCCGCGGTATGCTGACGCGCAGGTACCGCGCAACGCTGCGCGTG 7914
Db 1441 TGAACCTGCGGACCCGCGGTATGCTGACGCGCAGGTACCGCGCAACGCTGCGCGTG 1500
Qy 7915 ACCGAACAGGCGCTGACCGCGCTTCCGACGAGATGAGAGCGGTCAACCGCGGTGCGGACGCC 7974
Db 1501 ACCGAACAGGCGCTGACCGCGCTTCCGACGAGATGAGAGCGGTCAACCGCGGTGCGGACGCC 1560
Qy 7975 GCGGTCTGGTGGGGGTGTTGAGGCGGAGCTGAGCGGCGCAACGAGTGGGCAACGCGTCTG 8034
Db 1561 GCGGTCTGGTGGGGGTGTTGAGGCGGAGCTGAGCGGCGCAACGAGTGGGCAACGCGTCTG 1620
Qy 8035 GCCGTACCGAGCTTCCGAACTGAGCGGACGACTTTCGCGACGACCGCGAGACCGGACCC 8094
Db 1621 GCCGTACCGAGCTTCCGAACTGAGCGGACGACTTTCGCGACGACCGCGAGACCGGACCC 1680
Qy 8095 CGTGTGTGACGGAAGTGTGCGGACGAGTCCCGCGTCCACTTGAAAGCGCGCACCGCC 8154
Db 1681 CGTGTGTGACGGAAGTGTGCGGACGAGTCCCGCGTCCACTTGAAAGCGCGCACCGCC 1740
Qy 8155 GCGTGGACCGCGCGGTGCGGCGGTCGACGTCGCGACCGGTGCGGAGGTGACAGTGTGTC 8214
Db 1741 GCGTGGACCGCGCGGTGCGGCGGTCGACGTCGCGACCGGTGCGGAGGTGACAGTGTGTC 1800
Qy 8215 GTGCGCGCGCGAACCGTATCCGAGGCTTCAACCGATCCCGACCGGTTTCAGCTGAGAC 8274
Db 1801 GTGCGCGCGCGAACCGTATCCGAGGCTTCAACCGATCCCGACCGGTTTCAGCTGAGAC 1860
Qy 8275 CGTGCGCGGACGCGGAGATCTGTGTCGCGCGCGGCTGCGCTGCGCGCACCGACTGAC 8334
Db 1861 CGTGCGCGGACGCGGAGATCTGTGTCGCGCGCGGCTGCGCTGCGCGCACCGACTGAC 1920
Qy 8335 GCCCTGTGTGCGAACCTTGGCGACGCGCGGCTGCGCGCGCGCGCGCGGTTGCCCGG 8394
Db 1921 GCCCTGTGTGCGAACCTTGGCGACGCGCGGCTGCGCGCGCGCGCGCGGTTGCCCGG 1980
Qy 8395 CTGTCCCGTTCGCGGCGCGGTGATCAGACGACGTGCTCAACCGTCGCGGTGCTCAGC 8454
Db 1981 CTGTCCCGTTCGCGGCGCGGTGATCAGACGACGTGCTCAACCGTCGCGGTGCTCAGC 2040
Qy 8455 CTTTCCCGGTGAGCTGATGAGAGAAAGATGCGCGTCTGTGTTTCAATGATGAGCTGT 8514
Db 2041 CTTTCCCGGTGAGCTGATGAGAGAAAGATGCGCGTCTGTGTTTCAATGATGAGCTGT 2100
Qy 8515 CAACAGCCATCTGTTCCGCGCTGTCCCGTCCGACAGCGCTTCCAGCGCGCGGACAGCA 8574
Db 2101 CAACAGCCATCTGTTCCGCGCTGTCCCGTCCGACAGCGCTTCCAGCGCGCGGACAGCA 2160
Qy 8575 GGTACGGGTGCTGTGCGCTGCGCGGCTGACGACGTCACGAGGTCGCGGTCTGACCGC 8634
Db 2161 GGTACGGGTGCTGTGCGCTGCGCGGCTGACGACGTCACGAGGTCGCGGTCTGACCGC 2220
Qy 8635 CGTGCCGTGCGTGAACGCTGGAACCTTGTGAGGTGACGCGCCACGCGGCGCGAGCAT 8694
Db 2221 CGTGCCGTGCGTGAACGCTGGAACCTTGTGAGGTGACGCGCCACGCGGCGCGAGCAT 2280
Qy 8695 CGTGAGGTACATGCGGACCTGACGAGGTGTCACAGAGCCACACACATGCTCCGAGCA 8754
Db 2281 CGTGAGGTACATGCGGACCTGACGAGGTGTCACAGAGCCACACACATGCTCCGAGCA 2340
Qy 8755 CGACTCTCTGGGACATGACAGCACTTCAACCGCGACTTCTTCCGCTGATGAGCCCGCA 8814
Db 2341 CGACTCTCTGGGACATGACAGCACTTCAACCGCGACTTCTTCCGCTGATGAGCCCGCA 2400

QY 8815 CTCGCTCATGACGAGATGTCAGATTCTGCGGCTCTGCGGTCGCCAGCTGATCGTCTG 8874
DB 2401 CTCGCTCATGACGAGATGTCAGATTCTGCGGCTCTGCGGTCGCCAGCTGATCGTCTG 2460
QY 8875 GAGCCGCTGACCTTCCGCGCCCGGATCGCGGCGGTCACCGGAAACCCGCGACGCCCG 8934
DB 2461 GAGCCGCTGACCTTCCGCGCCCGGATCGCGGCGGTCACCGGAAACCCGCGACGCCCG 2520
QY 8935 GATGCTGTGGGGTCCGAGACGTGCGACCCCGGCGCGGATGCTTCCGTCGACGTCGTGC 8994
DB 2521 GATGCTGTGGGGTCCGAGACGTGCGACCCCGGCGCGGATGCTTCCGTCGACGTCGTGC 2580
QY 8995 CCACCGAGAGGTGAGCAGCCGAGAGATCC 9024
DB 2581 CCACCGAGAGGTGAGCAGCCGAGAGATCC 2610
RESULT 5
AA287284
ID AA287284 standard; DNA; 12441 BP.
AC AA287284;
AT 15-SBP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DB 5. venezuelae deosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
OS Streptomyces venezuelae; ATCC15439.
XX
PN MO20000620-A2.
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US014398.
XX
PR 26-JUN-1998; 98US-00105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
PI WPI; 2000-160679/14.
DR P-PSDB; AAY77179.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin.
PS
XX Claim 2; Page 281-287; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC streptomycetes antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragment thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthase may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to

CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster from *Streptomyces venezuelae* ATCC 15439.
CC (Updated on 15-SBP-2003 to standardise OS field)
XX
SQ Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 U; 0 Other;
QY
DB 2369 GGGCGGGTCGACGCGCGTTCAAGACGCGGATCAGGCGCCGATGATCCACCGCTCG 2428
DB 9472 GGGCGGATCGGGGCGGGTGCAGGCGCAGAGAACGCGGCGCTCTCCACGCGTCC 9531
QY 2429 GCGACGTCCTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2488
DB 9532 GCGCGCTGCGCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 9591
QY 2489 AAGAACTGTCCTCCGCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 2548
DB 9592 AAGTACTGTCCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 9651
QY 2549 ACCGAGTCAGAGCTCCGCTCGGATGATGATGATGATGATGATGATGATGATGATGATG 2608
DB 9652 ACTCGGTCAGAGAGTGTGCGGGGTGACCGCGCGCGATGATGATGATGATGATGATG 9711
QY 2609 AGCCCGGGAACCGCGCTCCGCTCAAGTATGATGATGATGATGATGATGATGATGATG 2668
DB 9712 AGGTGAGGAGACCGGCTCGGATGATGATGATGATGATGATGATGATGATGATGATG 9771
QY 2669 GCGACTGCGGATGAGGCGGTGCGGCGATGATGATGATGATGATGATGATGATGATG 2728
DB 9772 GCGACTGCGGATGAGGCGGTGCGGCGATGATGATGATGATGATGATGATGATGATG 9831
QY 2729 TCGACACCCGAGCGAGGCTCTCCACGCTGATGATGATGATGATGATGATGATGATG 2788
DB 9832 TCGGCGCGGAGCGAGGCTCTCCACGCTGATGATGATGATGATGATGATGATGATG 9891
QY 2789 CGCTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2848
DB 9892 CGCTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9951
QY 2849 TCGGACACCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2908
DB 9952 TCGGACACCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10011
QY 2909 GCGCGGTGCGGCTGAGACTGTTGATGATGATGATGATGATGATGATGATGATGATG 2968
DB 10012 GTCTGCGGCTGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 10071
QY 2969 TTGGCCCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028
DB 10072 GAGGACGCGCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 10131
QY 3029 GCGCGCATCGCGAGAGTCCCGAGGTTCTTCGACCCGTTGAGAGCGCGCGCTTG 3088
DB 10132 GCGCGCATCGCGAGAGTCCCGAGGTTCTTCGACCCGTTGAGAGCGCGCGCTTG 10191
QY 3089 CCGGTGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3148
DB 10192 CCGGTGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10251
QY 3149 AGGTTCACAGAGCGCGGTGCGCGCGAGCTGTTGATGATGATGATGATGATGATGATG 3208
DB 10252 AGGTCACAGAGCGCGGTGCGCGCGAGCTGTTGATGATGATGATGATGATGATGATG 10311

XX	Key	Location/Qualifiers
XX	CDS	3..809
FT		/*tag= a
FT		/product= "Streptomyces venezuelae Des gene cluster
FT		encoded protein #1"
FT		806..2014
FT	CDS	/*tag= b
FT		/product= "Streptomyces venezuelae Des VIII protein"
FT		/transl_except= (pos:806..808, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	2162..3442
FT		/*tag= c
FT		/product= "Streptomyces venezuelae Des VII protein"
FT		3535..4245
FT	CDS	/*tag= d
FT		/product= "Streptomyces venezuelae Des VI protein"
FT		/transl_except= (pos:3532..3534, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	4312..6741
FT		/*tag= e
FT		/product= "Streptomyces venezuelae Des R protein"
FT		/transl_except= (pos:4312..4314, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(7969..6829)
FT		/*tag= f
FT		/product= "Streptomyces venezuelae Des V protein"
FT		/transl_except= (pos:7969..7967, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(8979..7966)
FT		/*tag= g
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		/transl_except= (pos:8979..8977, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(9111..9989)
FT		/*tag= h
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		complement(11639..10182)
FT	CDS	/*tag= i
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		complement(12883..11636)
FT	CDS	/*tag= j
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		/transl_except= (pos:12881..12883, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
XX	PN	MO200229035-A2.
XX	PD	11-APR-2002.
XX	PE	05-OCT-2001; 2001WC-US031255.
XX	PR	05-OCT-2000; 2000US-0238185P.
XX	PA	(MINU) UNIV MINNESOTA.
XX	PA	(LIUH/) LIU H.
XX	PA	(SHSR/) SHERMAN D H.
XX	PA	(ZHAO/) ZHAO L.
XX	PI	Liu H, Sherman DH, Zhao L;
XX	XX	
XX	XX	WPI; 2002-405171/43.
XX	XX	P-PSDB; AAE24228, AAE24229, AAE24230, AAE24231, AAE24232, AAE24233,
XX	XX	AAE24234, AAE24235, AAE24236, AAE24237, AAE24347.
XX	XX	
XX	XX	Modified recombinant bacterial host cells in which the expression and
XX	XX	activity of nucleic acids encoding sugar biosynthetic enzymes has been

altered, useful for producing metabolites with altered sugar structures.

Disclosure; Fig 8; 174pp; English.

The invention provides a method to alter the sugar structure diversity for a particular metabolite via the recruitment and collaborative action of sugar genes from a variety of sugar biosynthetic pathways to yield a metabolite comprising a non-natural sugar, e.g., a novel glycosylated polypeptide. The invention also relates to a modified recombinant bacterial host cell (mBHC) in which the expression and activity of nucleic acid encoding sugar biosynthetic enzymes has been altered. The mBHCs may be cultured to produce the modified sugar products, e.g. a macrolide, anthracycline, amphotericin, avermectin, milbemycin, tetracycline, polyene, polyether, anemycin or isochromanequinone. The present sequence is Streptomyces venezuelae sugar (desosamine) biosynthetic gene cluster

Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 U; 0 Other;

Query Match 14.6%; Score 1309.4; DB 6; Length 13613;
Best Local Similarity 67.7%; Prid. No. 3,9e-170;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

2369 GGGGGGATCGACGCGCTTCAGACGCGCGGATAGAGCGCCGATGCTGATCCACCGCTGC 2428
10159 GGGGGGATCGGGGGCGGGTGGGGGTACGGCGCAGAAAGCCCGGGCTCTCCAGCGCTCC 10218
2429 GCGACGTCGCCGTTGCACTTGTTGAGCCCGGCGGTCACGCACTGTGTCAAAACGTTGAGG 2488
10219 GCGGCGTCGCGCTTCAGCTGTTCAAGCGGGCGGTGACACGACTGATCGAAAGCGTTCATG 10278
2489 AAGAACTGTCCTCCCGGCTCGGATGATGATCTGCGGCGCCGACGTACGAAAGTCGTGAGAG 2548
10279 AAGTACTGTCGCTCGCTTCGACGCGCGCACCTCCCGCGCTCGACGAAAGTCCCTGAGG 10338
2549 ACCGAGTCGAGGCTCCGGTCCGGGGTGAACCCCGCGCGGACGATGAGGAGGTCGCCCGCC 2608
10339 ACCTCGATGAGGAGGATGTGGGGGTGACCGCGGCGCGCATGTGAGCGAGTCCGCGCTTC 10398
2609 AGCCCGGAGAAACCGGCGCTCCCGGATCAGGTAGACGTCTTCGAGACAGTTCGACCC 2668
10339 AGCTCGGGAAACCGGCGCTCGCGGTACAGGTACGTCGCGGAGAGATCGACTCGACC 10458
2669 GCGACTGCGGATGAGGCGGTTCGGGCGCATCTGTTCCGGCGGATCCGCAACAGCTGAGCG 2728
10459 GCGACTGCGGATGAGGCGGTTCGGGCGCATCTGTTCCGGCGGATCCGCAACAGCTGAGCG 10518
2729 TCGACACCGCGGACGCAAGCTCTCGAACGCGTAAACCGAGTCGAGTCGACCGCGGAGTC 2788
10519 TCGGCGCGGTCGCAAGCTCTGTTGAGGCGTAAACCGAGTCGAGTCGAGTCGAGTCGAGTC 10578
2789 GCGTCGCGGTCGATGATGACGAACCGGCGCAGGCGCTTCGCGCAACTCGATCCGTTGCGCC 2848
10579 GCGTCGCGGTCGATGATGACGAACCGGCGCAGGCGCTTCGCGCAACTCGATCCGTTGCGCC 10638
2849 TCGGCAACCTGCGCTGCTCCGACCGCTGTAGTCTTCGCGGACGCGTACGAAGTCGAGC 2908
10639 TCGGCAACCTGCGCTGCTCCGACCGCTGTAGTCTTCGCGGACGCGTACGAAGTCGAGC 10698
2909 GCGCGGTCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2968
10699 GTCCTGCGCTGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10758
2969 TTGGCGCGCGCGCGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028
10759 GAGGCAAGCGCGCGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10818
3029 GCGGCAATCGGCAAGATCCCGGAGGTTCTTCGACCCGTTGAAAGCGCGCACGCTTG 3088
10819 GCGGCAATCGGCAAGATCCCGGAGGTTCTTCGACCCGTTGAAAGCGCGCACGCTTG 10878
3089 CCGGTGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3148

Db 10879 CCGGTGCTGCTGTACTCTCTGCTGTTAGAGCCGTAAGACGAGTGCAGATGGCGTGC 10938
 QY 3149 AGGCTCCACAGGCGGGCTGGCGCGGAGGCTCTGTTCCGTGAGGGGAGAGCGCTTGGTG 3208
 Db 10939 AGGCCCCAGAGGCGGGCTGGCGCTCTCAAGGATGCGCTCGGAGAGCGGAGAGGATTTGTG 10998
 QY 3209 TAGACGCTGAGGCGAGAGCCCGCGGCGCGGTGCGACACCAAGCCCGATACCGGG 3268
 Db 10999 TAGACGCTGAGGCGAGAGCCCGGTGCGTGGCCGCGCCGACAGCTCCGAGGCCGGGG 11058
 QY 3269 TTGGTCAACGGTTTCAGGCGCGCGAGAGGTACATGCGCCGTCCGTTCTCCAGCGAC 3328
 Db 11059 TTGGTGAAGCGCTTCAGGCGCGCGAGAGGTACATCCCGAGGGGTGGCCGCGGATATC 11118
 QY 3329 TCGTCAATGACCGAGGTAGGATCTGTTGCGCGGCGACCAACATCCGCGTCTATTTG 3388
 Db 11119 TCGTCAATGACCGAGCCGAAATGAGCTGTGCGCGCGGTGAGAGGAGCGGAGTCTTACCG 11178
 QY 3389 GCACTTGTCAACCGTACGAGAAAGTGAACAGCGGAAATGACATGACGTCGCGGCGGAGTAGC 3448
 Db 11179 GCGCCGCTCAACAGAGCGAGAAATGAGCAAGCAATGACAGTCCGAGCGGAGTTAGAG 11238
 QY 3449 CCGACGCTGTACGAGAAAGCGGAGCTCCGCGGTACCGACCTTCACACGCGCGGCGCG 3508
 Db 11239 CCGACGCTGTACGAGAAAGCGGAGCTTCGAGCGAGCGCGGTGAGAGACGCGCGCTGT 11298
 QY 3509 TTGAGGCGCTGATCCGTTGCTTCAGTACTTGGCCGCGCGGCGCTCTCTCAACGCGGTG 3568
 Db 11299 TCGAGCGGAGAGAGGAGGTTCTTTCAGTACGCGCGCGGCGGCGGTCTGACGCGCGGTG 11358
 QY 3569 CCGACGCTGCGGAGTGCATGCAAGGTGAGAGCTGCGCGAAACCGCGCGCGGTTCAGG 3628
 Db 11359 CCGAGCTTCGAGAGCTGCGCGAAACGAGAGAGCGCGGAGAGGCGTCCGCTGACG 11418
 QY 3629 TCGAACCGAGCGGCGATCTCTCAACGAGGAGTGAACGCGAGCGCGGTGAGTGGCGTG 3688
 Db 11419 CCGAGGCTGAGCGGAGCTCTCTCAAGCGGAGTGAAGGAGGCTGTTGCGGTGAGCGAGCG 11478
 QY 3689 AGTTGCGGAGCGGAGCGCGCTGCTGTTCCGCGGTGTCGATGCGGAGCGCGGTGAGCGG 3748
 Db 11479 AGCGGAGCGAGGTGCGGAGGCGGTGCTTCCGCGCTGTCGCGGCGCAAGAGCGCGCGCG 11538
 QY 3749 GTGACCTGCGGCGTGAAGCGCGCGGACCAAGCGCGCGCGGTGCGGCGTCCGAGCGCG 3808
 Db 11539 GCGAGGCTGAGCGCGAGCGGTGAGACCGCGCGCGCGGATCGGCTCCGAGGTGCGCAG 11598
 QY 3809 TCGCGCATCTCCGTGCGTACGAGCGGTGCGCGGTGAGTGGTCTTCAATGACGTGCGAAC 3868
 Db 11599 CGTTCCGCGCGGAGCGGTGCGGAAAGGCGGCGGCGGTCTATCGGAGCGTCCATCGTGG 11658
 QY 3869 CTTCGCGCTGCTGTGTCGAGAGATCAAGAACGTTGCGTTCCGCTTCCCATCTCCG 3928
 Db 11659 GCGTGAATGTCTGAGGCGCGCGAGCGGCGGCGCGGTGCGCGGAGTGC-----GCGC 11713
 QY 3929 GTTGATCAACGACCGCTGAGTGAAGTGAAGTCCGCAACCGCGGAGATGTC 3988
 Db 11714 GGTCACTGTCGCGCGCGGAGTCCGAGAGACGACGAGGTGAGCGACCGCGGAGATGTC 11773
 QY 3989 CTCACTGACACCGAGGAGCGCGGTGCGGAGAGCGAGACCGGTTTCCGAGCACTCGGT 4048
 Db 11774 GTCTGCGCGAGTGGCGGTGCGGTGCGGAGAGACGACGCGCGCGGAGGCGTTCCGT 11833
 QY 4049 CTGTGTACGCGGAGCGGTGCGGTCCGTGCGGTGAGGCGGTCTGTGTGAGCGGCGGAG 4108
 Db 11834 GTGCGGAGCGGCGGCGGTGCGGTGCGCGCGGTGCGGTGCGGTGCGAGCGCGCGAG 11893
 QY 4109 GAAGTGAAGCGGAGCGGAGCGGTTTCCGCGCGGAGATGCGTGAACCGCGGTCAAGGTC 4168
 Db 11894 GAAGTGAAGCGGAGGTGAGCGGCTTCCGCTTCAAGAGCTTCAATGACGAGTTCGCGGTG 11953
 QY 4169 GATGCGGAGCGGAGCGGTGCGGTCAACAGATGATCACTGATGATGCTCTCTCTGTC 4228
 Db 11954 GATGCGGAGGTGAGCGGTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 12013

QY 4229 GGGCGGAGAGGATGACAGGTGACGCGCTGCTAGTCCGCGACGCTCGCTGTGATGAGCCG 4288
 Db 12014 GCGGTGCTGCGGACGAGAGCGCGGAGAGTTCGCGAGGTGCTGCGGTGAGCGCG 12073
 QY 4289 GTGTTGACCCGTTGTGCTCCCTGCTTCGCGGAAAGCGCTGACGAGGAGGTGAGCCCAT 4348
 Db 12074 GTGTTGTCGCGGTTTCGCTGATGACCTCGGAAAGCGCTTCAAGGAGGTGAGGCCAT 12133
 QY 4349 GCGCGGCGGACCTGCTCATCTTTCGTTGTCGCGATCTGAGTCAACCTTGTCCG 4408
 Db 12134 GCGCGGCGGCGCTCCCTCATCTTTCGTTGTCGCGCGCGCGGCGGCTGCGCGGAG 12193
 QY 4409 GCGGATGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4468
 Db 12194 GTCGAGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12253
 QY 4469 CACGCGCGCGCTCGAAGCGGTGACCGGCTTGTGTCGTCGTCGTCGTCGTCGTCGTCG 4528
 Db 12254 GACGCGCGCGCTCGAAGCGGTGACCGGCTTGTGTCGTCGTCGTCGTCGTCGTCGTCG 12313
 QY 4529 GTCACCGGATTCGCGGACCGGAGCTGACCGGCTGTCGACCGGCGGCGGCGGCTG 4588
 Db 12314 GTCGCGAGGCTGCGCGGCGGCGGCGGCTGTCGACCGGCGGCGGCGGCGGCGGCTG 12373
 QY 4589 GAAGAAGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4648
 Db 12374 GAAGTACGCGGAGCGGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 12433
 QY 4649 TCGGCGCGCAGGTTACGCGGACGATGCGCGCGGTCTGCGGAGGTGACGAGCGCGGAC 4708
 Db 12434 GCGGCGCGGAGGTGAGCGGACGACGAGCGGTCGAGGTCGAGGTCGAGGTCGAGGTCG 12493
 QY 4709 GTGTCGCGGTCGACGAGACCGGTGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4768
 Db 12494 CTGTCGCGGTCGAGGTTGCGGTCGCGGTCGCGGTCGAGTCGTCGTCGTCGTCGTCGTCG 12553
 QY 4769 GAGCAGGCTGCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 4828
 Db 12554 GATCCAGCGAGTGTGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 12613
 QY 4829 GGTGACGTCACCGGCGCGGACGACGAGTTCAGAGGCGAGCGGTCGTCGTCGTCGTCGTCG 4888
 Db 12614 GGTGAGGCGGCGGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 12673
 QY 4889 GCACTGCGTACCGGACGAGTTCGCGGACCGGCGGTCGCGGTCGCGGTCGCGGTCGCGG 4948
 Db 12674 GCGATGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12733
 QY 4949 GCGGTCGTCGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5008
 Db 12734 GCGGTCGTCGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 12793
 QY 5009 GTGCGGATGTCGCGGCGGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5068
 Db 12794 GTGATGCGGTCGTCGCGGCGGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 12853
 QY 5069 GATCGCAGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5099
 Db 12854 GAATGCGAGTTCGAGTAAGCGGCTTTTACG 12884

RESULT 7
 ADL91931
 ID ADL91931 standard; DNA; 13613 BP.
 XX ADL91931;
 AC
 XX 20-MAY-2004 (first entry)
 DT
 XX Streptomyces venezuelae desosamine gene cluster coding sequence.
 DE
 XX narbonolide polyketide synthase gene; polyhydroxyalkanoate monomer;
 KM

XX 06-JAN-2000.
PD
XX
XX 25-JUN-1999; 99MO-US014398.
PF
XX 26-JUN-1998; 98US-00105537.
PR
XX
XX (MINU) UNIV MINNESOTA.
PA
XX Sherman DH, Liu H, Xue Y, Zhao L,
PI
XX
XX WPI; 2000-160679/14.
DR
XX P-BSID: AAY77204, AAY77205, AAY77206, AAY77207, AAY77208, AAY77209,
DR AAY77210, AAY77211, AAY77212, AAY80998, AAY80999.
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin.
PS
XX Diaclosure; Fig 32; 438bp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC streptomycetes antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
CC pikromycin, neomethymycin, nadomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC nadomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439,
CC as given in figure 32. (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 U; 0 Other;
Query Match 14.6%; Score 1307.8; DB 3; Length 13613;
Best Local Similarity 67.7%; Pred. No. 6.4e-170;
Matches 1849; Conservative 0; Mismatches 877; Indels 5; Gaps 1;
QY 2369 GGGCGGCTGACGCGGCTTCAGACGGCGGATCAGAGCCCGATGCTATCCACCCGCG 2428
DB 10159 GGGGGGATCGGGGCGGGGTGCGGGTCAAGCGAGAGCCGGGCTCTCCAGCGCTCC 10218
QY 2429 GCGAGCTCCGTTGATGATTGTTAGCGCGGGCGGTACACGATGCTGAAACCGTGAAG 2488
DB 10219 GCGGGGTGCGGCTTCAGCTGGTTCAAGCGGGCGGTGACGACTGATCAAGACCGTCAATG 10278
QY 2489 AAGAACTGTCTCCCGCGCTGCGGATGATGCTGCGGCGCGAGCTGACGAAATGCTGAGACG 2548
DB 10279 AAGTACTGTGCGCCGTTCAGACGCGCCACCTTCGCGCGCGCTTCACAGAAAGTCCCTGACG 10338
QY 2549 ACCGAGTGAAGGCTCGGATCGCGGATGACCGGCGCGGAGCGTACGAGGATGAGCGCGCGC 2608
DB 10339 ACCTCGGTAGGAGAGATGTCGAGGATCACGCGGCGCGGATTAAGGGGTGCGCGCTCC 10338
QY 2609 AGCCCGGAGAAACCGGCGCTCCCGGTACAGGTAGACGTCTCCGAGCAGGTGATCTGCAAC 2668
DB 10399 AGGTGCGGAGAACCGGCGCTCCGGGTACAGGTACAGTCCGCGAGGAGATCGACTGCAAC 10458

QY 2669 GCCACCTGGAGATGAGGCGGATCGGGCGCATGTTCCGCGCCGATTCGCAACAGCTGCGCG 2728
DB 10459 GCGACCTGCGCGCTGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 10518
QY 2729 TCGACACCCCGACGACGAGGCTCTCCAAAGCGGTAAACCAAGGTGCTGATGATGCGGGGTC 2788
DB 10519 TCGGCGCGGATGCGAGGCTGTTCAAGGCGCTTACCGGTAGTGTGTGAGTATCCGGGGGTG 10578
QY 2789 CGCTCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2848
DB 10579 CGCTCGCGGACCGGCTCTCCAAAGCGGTAAAGGCGCTCTGAGAGCTGAGCGCGCTCTCC 10638
QY 2849 TCGGACAACTGCGCGCTGCTCCGACGCGCTGATGCTCTCGCGGACGCTGACGAAATCGAGC 2908
DB 10639 TCGGCGAGCTTGCCTGCTGATCAAGCGCGCTGATGCTCTCGGAAATGTTAGAGATGATC 10698
QY 2909 GCGCGGTGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2968
DB 10699 GTCTGCGCTGCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 10758
QY 2969 TTGGCGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028
DB 10759 GAGGACAGGCTCGGAGACAGATGTAGGCGAAGCGGATTTAGTTCGCGACTGCGCTTCG 10818
QY 3029 GCGGCGATTCGCGAGAAAGTCCCGAGAGTCTTCTCGACCGGTTTCGAAAGCGCGCACGCTTG 3088
DB 10819 GCGGCGAGCTGCTGAGAAACCGCGCGCGATTTCTCGCGACGCGGCGAAAGCGCGCTTCTTG 10878
QY 3089 CCGGTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3148
DB 10879 CCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10938
QY 3149 AGGTCCCAACGCGCGGCTGCG 3208
DB 10939 AGGCGCCAGAGCGCGGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10998
QY 3209 TAGACGATGAGGCGGAGAACCG 3268
DB 10999 TAGACGATGAGGCGGAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11058
QY 3269 TTGATCAACGATTCAGGCG 3328
DB 11059 TTGATGAGCGGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11118
QY 3329 TCGTGTATGACCGAGGTGAGGATCTCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3388
DB 11119 TCGTGTATGACCGAGGATGAGGATCTCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11178
QY 3389 GCACCTGTCAACCGTATGAGAGAGTGAACGCGGAAATGACATGCTCGCGCGCGCGCGTGAAGC 3448
DB 11179 GCGCGGATGACAGGACCGAGAAATGCGAGGAAATGCGAGTTCGCGCGCGCGCGTGAAGG 11238
QY 3449 CCGACGCTGTACGAGGAGCGGCGCTCCCGCGGTACCGGACGCTCCAAACGCGCGCGCGCG 3508
DB 11239 CCGACGCTGTACGAGGAGCGGCGCTCCCGCGGTACCGGACGCTCCAAACGCGCGCGCGCG 11298
QY 3509 TTCAAGCGCGCTGATGCTGTTGCTCAATGATCTGCGCGCGCGCGCGCGCGCGCGCGCGTG 3568
DB 11299 TCGAGCGGAGACAGGAGGTTCTTCAGATACCGCCCGCGCGCGCGCGCGCGCGCGTG 11358
QY 3569 CCGAGCTCGGAGATGAGATGACAGAGGTGACAGAGCTTCGCGGAAACCGCGCGCGCGTTCAGG 3628
DB 11359 CCGAGCTTCGAGGACCTGCGCGGAAACAGGCGGAGAGCGCGCGGAAAGCGTCCCGGTTCAGG 11418
QY 3629 TCGAACCGACGCGCATCTGCTCCACAGGAGTGAACCGGAGTGAACCGCGCGCGTGAAGG 3688
DB 11419 CCGAGGTGTGAGCGGCGCTCTCCAGCGCGGAGTGAAGGAGGCTGTTGCTGTACGCGACGCG 11478
QY 3689 AGTTCGAGAGACCGGACCGCTGCTGCTTCGCGCGGTTCGAGTTCGAGTTCGAGTTCGAGTTC 3748
DB 11479 AGCGGAGAGAGTGTGCGGCGGTGCTGCTTCGCGCGGTTCGAGTTCGAGTTCGAGTTCGAG 11538

[illegible]

Accession	Gene	Protein	Function	Location/Qualifiers
D8	12614	GATTAGGCGCGGCGGTCGCGAGAGCTTGAGCCCGGCGCGTGGCGTT		
OY	4889	GCAAGTGGCCTACCCCGACCAAGTCCGCGACCCCGGCGCTCGAATCTCCCG		
D8	12674	GGCATTGCGGACCCCGGCGAGCGCCGCGACGCTCTCTCGAATCTCGCC		
OY	4949	GCCGTGGTGAACCGATTGTTGTCAGGGCCCACTCGAGCGCGGCCAA		
D8	12734	GCCGTTGGACAGCCACTGCTGTGAGGGGCCCGGTGAGCGCGCTGTT		
OY	5009	GTCGCCGATCGTCGCGCGGCCCGACGTGCAAGGGGGTGCAGAAACGCTT		
D8	12794	GTCGATGGCGGTTGGGCGCGCCCGACGAGAGCGGCTGTGGAAGCGGG		
OY	5069	GATGCGCAGATCCGATCGGTCGAGCGCGCTTCAAG 5099		
D8	12854	GAATCGAGTGTGCGATTAAGCGGCTTTTTCACG 12884		
RESULT 9				
AAD55810/c				
ID	AAD55810	standard; DNA; 60196 BP.		
XX	AAD55810;			
XX	27-OCT-2003 (revised)			
DT	07-AUG-2003 (first entry)			
XX				
DE	Micromonospora carbonacea polyketide synthase (PKS) type			
XX				
KM	Macrolide; rosaramicin; polyketide; polyketide synthase;			
XX	gene; ds.			
OS	Micromonospora carbonacea.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	1..1683		
FT		/*tag= a		
FT		/product= "polyketide synthase #1"		
FT		/note= "CDS does not include start codon"		
FT		/partial		
FT	CDS	complement(1728..2522)		
FT		/*tag= b		
FT		/product= "polyketide synthase #2"		
FT		complement(2629..3861)		
FT	CDS	/*tag= c		
FT		/product= "polyketide synthase #3"		
FT		4365..5573		
FT		/*tag= d		
FT		/product= "polyketide synthase #4"		
FT		5702..19117		
FT		/*tag= e		
FT		/product= "polyketide synthase #5"		
FT	CDS	19144..24921		
FT		/*tag= f		
FT		/product= "polyketide synthase #6"		
FT		/note= "CDS does not include start codon"		
FT		/partial		
FT	CDS	24993..36230		
FT		/*tag= g		
FT		/product= "polyketide synthase #7"		
FT		36292..41016		
FT	CDS	/*tag= h		
FT		/product= "polyketide synthase #8"		
FT		41049..46403		
FT		/*tag= i		
FT		/product= "polyketide synthase #9"		
FT		46400..47794		
FT		/*tag= j		
FT		/product= "polyketide synthase #10"		
FT		/note= "CDS does not include start codon"		
FT		/partial		

ID ADM45913 standard; DNA; 84428 BP.
XX
AC ADM45913;
XX
DT 03-JUN-2004 (first entry)
XX
DE Streptomyces mycarofaciens midcamycin polyketide synthetase DNA.
XX
KM macrolide; midcamycin biosynthesis; polyketide synthetase; PKS;
XX antimicrobial; ds; gene.
OS Streptomyces mycarofaciens.
XX
PN JP2004049100-A.
XX
PD 19-FEB-2004.
XX
PF 19-JUL-2002; 2002JP-00210516.
XX
PR 19-JUL-2002; 2002JP-00210516.
XX
PA (MEIJU) MEIJU SEIKA KAISHA LTD.
XX
DR WPI; 2004-174115/17.
XX
DR P-PSDB; ADM45914, ADM45915, ADM45916, ADM45917, ADM45918, ADM45919,
XX ADM45920, ADM45921, ADM45922, ADM45923, ADM45924, ADM45925, ADM45926,
XX ADM45927, ADM45928, ADM45929, ADM45930, ADM45931, ADM45932, ADM45933,
XX ADM45934, ADM45935, ADM45936, ADM45937, ADM45938, ADM45939, ADM45940,
XX ADM45941, ADM45942, ADM45943, ADM45944, ADM45945, ADM45946, ADM45947,
XX ADM45948, ADM45949, ADM45950, ADM45951, ADM45952, ADM45953, ADM45954,
XX ADM45955.
XX
PT New polynucleotides encoding proteins important in the biosynthesis of
PT midcamycin.
XX
PS Claim 2; SEQ ID NO 1; 264bp; Japanese.
XX
CC The invention relates to a novel polynucleotide, specifically a
CC biosynthesis gene, comprising a nucleotide sequence which encodes a
CC protein, where the gene is concerned with the biosynthesis of
CC midcamycin. The polynucleotide of the invention demonstrates a
CC antimicrobial activity and may be useful for the biosynthesis of
CC midcamycin, as well as for manufacturing macrolide compounds other than
CC midcamycin. The current sequence is that of the Streptomyces
CC mycarofaciens midcamycin polyketide synthetase (PKS) DNA of the
CC invention.
XX
SQ Sequence 84428 BP; 11456 A; 29098 C; 30275 G; 13599 T; 0 U; 0 Other;
Query Match 13.5%; Score 1203.2; DB 12; Length 84428;
Best Local Similarity 65.5%; Pred. No. 1e-155;
Matches 1778; Conservative 0; Mismatches 933; Indels 5; Gaps 1;
QY 2387 CAGACGGCGCGATGATGAGCCCGGATGATCCACCCGTCGGGACGTCCTCGATG 2446
DB |||||
QY 3232 CAGGCGTATCAAGAAAGCCGCGTGTGTCGCCCAACCGTCGGCGGATTCATCTCGAGC 3291
DB |||||
QY 2447 TGGTTGAGCCCGGCGGTCAACGACTGTGTGAAACCGTGAAGAAACTCGTCCCGCGGC 2506
DB |||||
QY 3292 TGGTTCAACCGTGTGATGACGACTGTGTGAAACCGTCAATGAATCTGTGCGCTCG 3351
DB |||||
QY 2507 TGGCGATGATGCTGCGGCGCCGACGTCGACGAAGTCGTGAACGACGAGTGCAGCTCCGG 2566
DB |||||
QY 3352 GCCGAGGAATCTGTCCCAACCGCTGTCAAAAGCGCTCAACCTGGGTCAAGCTTTT 3411
DB |||||
QY 2567 TCCGGGATGACCCGCGCCGACGATGACGGGTGACCCCGCCGCGGAGAAACCGGCGC 2626
DB |||||
QY 3412 CCGGGGAGACCTTGCCGATGTGTAGCGGTCCGCTCCGGGACGTCGCGGAAACCGGCGC 3471
DB |||||
QY 2627 TCCGGTACAGGTACAGTCTTCCGAGCAGTGTGATCTGACCGCCACCTTCGGGATGGCGC 2686
DB |||||
QY 3472 TCGCGTAGAGGTAGACGTGCGCCAGCAGATCGACTGTGACCTGCGAGTGGCGCG 3531
DB |||||

QY 2687 GTCGGGCGATCGTTTCGCGCGGATCCGCAACGCTGGCGTTCGACACCCCGACGCAAG 2746
DB |||||
QY 3532 GTCGGCGCATGTGTTTCGCGCCGAATCCGACGAGTTTCGCGCTCCGCGCGGACG 3591
DB |||||
QY 2747 CTCTCCAAACCGGTACCCCAAGTGGTCTGTGATCCCGGGGTCCGCTCGGCGGTAGTCG 2806
DB |||||
QY 3592 CTGTTCAACCGGTACCCGATGTCGACGAGGTGGTGTCTCTGCGCGGACCTTCTCC 3651
DB |||||
QY 2807 ACGAACCGGCGGAGCCCTCCCGCACTCGATCGTTCGCTTCGACCAACTGCGCTCG 2866
DB |||||
QY 3652 CGAAAGCGGTCAAGCGCTTCTGTGAGGTCCGCGGCTCGGACCGGACGAGGCGGTG 3711
DB |||||
QY 2867 TCCGACCGGTGATGCTCTCGCGACGCTGACGAACTGACGCGGCGGTGCGGCTGAC 2926
DB |||||
QY 3712 GCGCGCGCGTGTAGTCTCAACGAGGTGACGAACTGATGATCGATCGCTCGCGCGCT 3771
DB |||||
QY 2927 TCGTTGAGTTTCGCGGATGAAAGTGCACGAGTTCGATGAGGCGTTTGGCCCGCGGAG 2986
DB |||||
QY 3772 GCGTTCAACTCGGAGTAAAGTGCACGAGGTTCGAGCAGGCGCGGCCCGGACCGGACG 3831
DB |||||
QY 2987 ATGATGTGTTGAGGCGGACCGGACCGGTGCGCCCGTTTCGCGCGCATTCGACGAGAG 3046
DB |||||
QY 3832 ACGATGTAGTTCAAGCGCGGAGTTTCAACGGTTTCGCGCGGTGCGGACCGCGTGAAG 3891
DB |||||
QY 3047 TCCCGAGGTTCTTCTGACCCGTTTCAAGGCGCCACGCTTCCGCGGTGCGCTGATC 3106
DB |||||
QY 3892 CGCTCCAGGTTGCGCTTCAACCGGTCCGAAAGCGGCTTCTTCTGTTGTTGCGCGTAT 3951
DB |||||
QY 3107 TCGTCTGTTTTCAGGCGCGGTCAAGGAGTTCGAGCGCGCGGAGTTCGACGCGCGG 3166
DB |||||
QY 3952 TCGCATGTTCACCCGTTACAGGAGTTCGACCGCGTTCAGGCGCGGACCGGACCGCC 4011
DB |||||
QY 3167 TGGCGGCGCAGCGTCTGTTGCTGAGGCGGAGCGTGTGTGACGCTTCAGGCGGAG 3226
DB |||||
QY 4012 TGCTTTTCAGCGTCCGCTCGGTACGAGGCGGAGCGTGTGTACAGGTAAGCGGAAAT 4071
DB |||||
QY 3227 CCGCGCGCGGCGGCGGTGACCAACGACCGCCGATCCCGGTTGTTCACGTTCCAGG 3286
DB |||||
QY 4072 CCGCGGCGCGGCGGCTTCCGAAACGAGGAGCCGCGCGGATGTGTGAGCGGCTCCGAG 4131
DB |||||
QY 3287 CCGCGGAGAGTATCATCGCGGTGCTTCGACCGGACCTGTGTGATGACGAGGATG 3346
DB |||||
QY 4132 CCGGCGGACAGTACATGCGCTCGGAGTGTCTGTTGACATGTGATGATGAGCGGCG 4191
DB |||||
QY 3347 AGATCTGTTGCGGCGACACCGACTCCGCTGTACTTGTGACACTGTACCCGTAAG 3406
DB |||||
QY 4192 AACATCTTGTGCGGTGTCCAGGCGGAGTGTGTAGCGGCGGCTGATCCGTAAG 4251
DB |||||
QY 3407 CAGAAGTGACAGCGGAAATGACATGTCGCGGCGGAGTGAAGCCGCTGTACGGAAG 3466
DB |||||
QY 4252 CAGAAGTGACAGCGGAAATGACATGTCGCGGCGGAGTGAAGCCGCTGTACGGAAG 4311
DB |||||
QY 3467 GCGGCGTCCGCGCTACCGGACCTTCAACGCGCGCGGCGGTTCAAGCGGCTGATGCTG 3526
DB |||||
QY 4312 ACCGCTTGTGTGCAAGCGGCTGTGAAACGCTTTCGTTTCAGCGGAGGAGGAGT 4371
DB |||||
QY 3527 TTGCTCAAGTACTTTCGCGCGGCGCTGTCTCAACCGCGCTGACGCTGCGGATGCGA 3586
DB |||||
QY 4372 TTGCTCAAGTACTTTCGCGGAGGCGGCTGTCTCAACCGCGGAGGAGTCAAGGAG 4431
DB |||||
QY 3587 TCGAACAGGTGACAGCTTCGCGAACCCGCGCGGTTCAGGTGAACCGACGCGCATC 3646
DB |||||
QY 4432 CCGAAAGGTTCACAGACGCGCGGAGGACAGATGCTGTCCTGATTGTATGACGCGAC 4491
DB |||||
QY 3647 TGCTTCAACGCGGATGAAACGAGAGCGCGGTGATGTGGCGGTGAGTTCCGCGGACG 3706
DB |||||
QY 4492 TTTTCAGCGGAGTGAAGGAGCGCGCGCTGATGATCCGAGCAAGCTGATGAGCG 4551
DB |||||
QY 3707 GCGGTGCTTTCGCGGATGATGAGTGTGCGGACCGGATGAGGAGGATGACTTCGCGCGT 3766
DB |||||
QY 4552 GCGGCTTCGCGGATGATGCGGCTGCTGATATGCGCGGAGATGACAGTCTTTCGCGGAG 4611
DB |||||
QY 3767 GCGGCGACCAACGCGCGGCGGTGCGGCTGCGGCTGTGCGCGCTCGGATCTCGGT 3826
DB |||||

```

Db      4612 GCCTCGAAGCGGCGCGGATCACTCGCGAAGGGGTGCAAGCCGCTACGGTGTGCGC 4671
Qy      3827 ACCGCGGTGCGCGGAGAGTGTCTTATGACGCTGTGGAACCCCTTCTGGCGCTCTGTGTG 3886
Db      4672 ACCGCTTCTTGGCTCACCATGATGATCACTCCACCTCTTCCATTGCGAAGATGCTTGTGATG 4731
Qy      3887 CGAGATCA-----CGAAGCGTTGCGTTTCCGCTTGTCCCATCCCGGTTGATCAAGCA 3941
Db      4732 AGAAGGCGCGTCCGCTGTGACTGAGTGGCCCTGAGCGGCGGCGGTGACGTCCGCC 4791
Qy      3942 CCGTGTGTGTGCGAGTGTGATGACGTGCAACCCGCGGATGTCTCACTGAGAAC 4001
Db      4792 CCGGCGGTGCGCGCAACCGGACGATGTCCGATACCGCGGAGATATCTCATGGAAGAC 4851
Qy      4002 GAGGCGCGGTGCGGAGGAGGACCCGTTCCGCGAGCCACTGGTCTGTGTACGCCG 4061
Db      4852 GTGACCCGGTTCGCGAAGCGATGACCCGCGCGAAGCGCTCGGTGTGCGGAGGTGC 4911
Qy      4062 AGCGGTGTCTCCGTGCGGTAGAGGCGTATGTGTGAGGCGGAGGAGAGTAGGCTGT 4121
Db      4912 GCGTCCGAGGAGACCGGTACGGTCCAGTTGGTGGAGCGCGCGAAGAGTAGGCTGG 4971
Qy      4122 GCGACGACCTTCTCCGCGCGAGAGTGCCTGACCGCGTCACGGTGATGCGGTGAG 4181
Db      4972 GCCACGACGTTTTCGCGCGAAGGAGGAGGAGGAGGAGATGCGATGAGCGCGGTGAC 5031
Qy      4182 GTGCCGTTCACAGAGATGATCACTGATGTGTCTCTCTGTGTGCGGCGGAGCGAG 4241
Db      5032 TCTCTGTGATCTGACAGACAGATGATGTGTCTCTGTGTCTCATTCGAATGCG 5091
Qy      4242 TGCACGATGACGCGCGGTACGTCGCGAGCTCGTGTGTGTGAGCGGTGTGACCGCG 4301
Db      5092 AAGACGCGCACCGCGGTACTCCGAGAGTTTCGTGCGGTGATGCTCATGTGTCTTTG 5151
Qy      4302 TTGTGCTCCCTGTCTTCGCGAAGCGCTGAGGAGGTGAGCCCATGCGCGCGCAC 4361
Db      5152 TTGTGCGGTACGACTTCTCAACAGCTGAGGAGGAGTCAAGCCCATGCGCGAGGCT 5211
Qy      4362 TCGCTCATCTTGGCGGTGTGTGCGGATCTCGGTACCACTTGTGCGGCGCGATGCGAG 4421
Db      5212 TCGCTCATCTTGGCGGTGTGTGCGGCGGCGGAGTGACTTCTTCCAGCGCGAGTCCAG 5271
Qy      4422 TTGTGATGCGCGGATCGTGTGCGGAGGCGCGTGTGTGATCAACCGCGCGCGC 4481
Db      5272 TTGTGAGAGAGGAGGAGACATGAGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 5331
Qy      4482 TCGAAGCGGTGACCGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4541
Db      5332 TCGAAGGATGACGACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5391
Qy      4542 CCCACCGGACGTCAACCGTGTGCAAGCCGAGGCGGTGTGTGTGTGTGTGTGTGTGTGT 4601
Db      5392 CCCACGAGTCTTCTGTGCGAGGTGTGCGGATGCGGCGCGGTGTGTGTGTGTGTGTGT 5451
Qy      4602 ACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4661
Db      5452 ATGCCGCGCTTCCGCGGCAAGCTTCTTCAAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 5511
Qy      4662 TGTACGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4721
Db      5512 TGTACCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5571
Qy      4722 ACCAAGCCGTCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4781
Db      5572 ACGACATCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5631
Qy      4782 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4841
Db      5632 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5691
Qy      4842 GCCCGACGACGATTCAGGCGGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4901

```

```

Db      5692 GCGTGTACAGATGATTGAGCGGACGAGCGTGCATGTGAGGTGCGCACGACGATTGCGGACC 5751
Qy      4902 CCGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4961
Db      5752 CCGCGCAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5811
Qy      4962 CAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5021
Db      5812 CACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5871
Qy      5022 GCGCGCGCCACGTGACAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5081
Db      5872 GCGCGCGCCACGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5931
Qy      5082 GTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5097
Db      5932 CCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5947

RESULT 11
AAA75638/c
ID AAA75638 standard; DNA; 1565 BP.
XX
AC AAA75638;
XX
DT 22-JUN-2001 (first entry)
XX
DE Nucleotide sequence of the ORF of the picCV gene.
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KW desosaminyl transferase; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PE 27-MAY-1999; 99US-00320878.
XX
PR 30-APR-1997; 97US-00846247.
PR 06-MAY-1998; 98US-00073538.
PR 28-MAY-1998; 98US-0087080P.
PR 28-AUG-1998; 98US-00141808.
PR 22-SEP-1998; 98US-0100880P.
PR 08-FEB-1999; 99US-0119139P.
PR 20-MAY-1999; 99US-0134990P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
Ahley G, Betlach MC, Betlach M, Tang L, McDaniel R;
WPI; 2000-610844/58.
XX
DR
XX
PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value.
XX
PS Disclosure; Col 45-48; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compound
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode a
CC C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity of
CC a compound relative to the unhydroxylated compound. The recombinant host

```

CC cells are useful as genetic systems that allow rapid engineering of the
 CC narbonolide polyketide synthase. These would be valuable for creating
 CC novel ketolide analogs for pharmaceutical applications

XX Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;

Query Match 8.2%; Score 729.4; DB 3; Length 1565;
 Best Local Similarity 68.1%; Pred. No. 7.5e-91;
 Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

2369 GGGCGGCTGACAGCGCTTTCAGACCGCGCGATCAAGCGCCCGATCTGATCCACCGCTG 2428
 1530 GGGGGGATCGGGGCGGGTGCAGCGCGAGAAAGCGCGGCTCTCCCTCCAGCCCTGTC 1471
 2429 GCGACGCTCCGCTTTCAGATTGGTTGAGCGCGCGCTCAACGATGTGAAACCGTCGAGG 2488
 1470 GCGCGCTGCGCTCACTGATGTTGAGCGCGGGGTGACACTGATGAAAGCCCTTCATG 1411
 2489 AAGAACTGTCCTCCGCGCTGCGATCGATGCTGCGCGCCGACGTGAGAAAGTCGTGACG 2548
 1410 AAGTAATCTGTCGCTGCGATCGACCGCGCACCTGCGCGCGCTGACGAGATCCCTGACG 1351
 2549 ACCGATGTCAGAGCTCCGCTCGGGGTGACCGCGCGGACGTAGCGGGTGGCCCGCC 2608
 1350 ACTCGGTGAGGAGGTGTGCGGGGTGACCGCGCGCGATGTAGCGGGTGCCTGCTCC 1291
 2609 AGCCCGGGGAAACCGCGCTCCCGGTACAGGTAGACGTCTCCGACAGGTGATGACG 2668
 1290 AGGTGAGGAAAGCGGCTCGCGGTACAGGTACAGTCCGAGAGATTCACCTGACAC 1231
 2669 GCCACTGTCGAGTGGCGGTGCGGCGCATCTGTTCCGCGCGGATCCGACAGCTGAGCG 2728
 1230 GCGACTGTGCGGTGCGGGTGGCGCATGTTGGCGGGCTTGATCCGACAGTTCCGCG 1171
 2729 TCGAACCCCGACGAGGCTCTCAACCGGTACCCAGGTGCTGTGATCCCGGGGTC 2788
 1170 TCGGCGCGGTGCGCAGGCTGTTCAAGGCGGTAGCGGTGATGTGAGTCCGGGGTGG 1111
 2789 CGCTCGGCGGTGATGTCAGAACCGGCGAGGCGCTCCGCACTCGGTTCGGCC 2848
 1110 CGCTCGGCGAGCCCGCTCTCTCAAGAGCGTTGAGGCGCTCTGAGAGTCCGCGCTCTCC 1051
 2849 TCGGACAACTGCGCTGCTCCGACCGGCTGATGCTTCGCGGACGAGTCAAGTCAGC 2908
 1050 TCGCGAGACTTGCCTGTCACAGCGCGCTGATGCTTCGCAATGTTGACAAAGTCATC 991
 2909 GCGCGGTGCGGCTGACTGCTGTTGAGTTCGCGGATGAATGACCAAGTTCATGAGCGG 2968
 990 GTCTGCGCTGCGCGCTGCTGTTGAGTTCGCGGATGAATGACCAAGTTCAGAGCGG 931
 2969 TTGCGCGCGCGCGGAGATGATGTTGAGGCGGACCGGACCGGTCGCGCGCTTCG 3028
 930 GAGGACCGCGCGGAGACATGATGAGGCGGAAAGCGGATTGATCGGACCTCGGCTCG 871
 3029 GCGCGCATCCGACGAGATGCTCCGAGGTTCTTCTGACCCGTTCAAGAGCGCCACGCTTG 3088
 870 GCGCGCACTGCTGAAACGCGCGCAGTTCCTCGCGAGCGCGCGGAAAGCGCGCTTCTTG 811
 3089 CCGGTGCTGCGCTGCTGCTGCTGTTGACGCGGTACAGGAGGTACGAGCGCGCG 3148
 810 CCGGTGCTGCTGCTGCTGCTGCTGTTGAGGCGGTGAGGAGGTGCGGATGCGCTG 751
 3149 AGGTCCCAACAGCGCGCGGCTGCGCGGCGGCTGTTGAGGAGGCGGAGCGGCTTGCTG 3208
 750 AGGCGCCAGAGCGCGGCTGCGGCTGCGAGGCGCTGCTGAGGCGGAGGAGTTCTG 691
 3209 TAGACGCTGAGGCGGAGCGCGCGCGGCGGCTGACACCAAGCGCCGATACCGCGG 3268
 690 TAGACGCTGAGGCGGAGCGCGCGGCTGCTGCTGCGGCGGAGCTCCGAGGCGCGG 631
 3269 TTGCTCAACGCTTTCAGGCGCGCGGAGGATGATGCTGCTGCGGCTTCTCGACCGCAC 3328
 630 TTGCTGAGCGGCTTCAGGCGCGCGGAGGATGATGCTGCTGAGGAGGCTTCCCGGGTATC 571

3329 TCGTCATGACCGAGGTGAGGATCTGTTGCGCGGACCAACGATCTCCGCTGCTGTTG 3388
 570 TCGTCATGACCGAGCGGAAATGCGTTTCGCGCGCTGAGGCGGACGCGGTGATGCGG 511
 3389 GCACTGTGACCGGCTGACGAGAAAGTGAACAGCGGAACTGACGCTCGGCGCGGCTGAGC 3448
 510 GCGCCGCTGACAGGAGCGGAAAGTGAACAGCGGAACTGACGCTCGGCGCGGCTGAGC 451
 3449 CCGACGCTGACAGGAGCGGCGCTCCGCGCTGACGCGACCTTCAACGCGCGCGCG 3508
 450 CCGACGCTGACAGGAGCGGCGCTTCTGCGAGCGCGCGCTGAGAGAGCGCGCGCTG 391
 3509 TTGACGCGCTGATGCTGTTGCTCACTGCTGCGCGCGCGCGCTGCTCAACGCGCTG 3568
 390 TCGAGCGGAGGAGGAGGCTTCTTCACTGACCGCGCGCGCGCGCTGCTGACCGCGCTG 331
 3569 CCGAGCTGCGGAGTTCGATGAAACAGGTGACGAGCTGCGGAAACCGCGCGCTGCTCAAG 3628
 330 CCGAGCTTCCGAGACTGCTCCGAAACAGGCGAGGAGCGCGGAAAGCGCTCCCGGTGACG 271
 3629 TCGAACCGAGGCGCATCTGCTCAACGCGGAGTGAACGCGAGCGCGCTGATGAGCGG 3688
 270 CCGAGCTGCTGAGCGGCTCTTCAACGCGGAGTGAAGGAGCTGTTGCCGTGACGACGCG 211
 3689 AGTTGCGCGAGCGGACCGCGCTGCTGCTGCGCGGTGTGAGTGCAGCACCGGTGAGCGG 3748
 210 AGCGGACGAGGAGTGGCGGCGGTGCTTCCGCGCTGCTGCGGCGGACGAGGCGCGCGG 151
 3749 GTGACTTGGCGCTGAGAGCGCGCGACCAAGCGCGCGCGGTGCTGCGGTGCTGCGCGG 3808
 150 GCGAGGCTGAGCGCGAGGCGGTGACCGCGCGCGCGCGCGCGCTGCGGAGTGCAGCAG 91
 3809 TCGGAGTCTCCGCGGATGAGCGCGGTGAGGAGTGTCTTCATGAC 3859
 90 CGTTGCGCGGCGGCGGTGCGAGAGGCGCGGCGCGGTTCATCGGAGCGCT 40

RESULT 12
 AA25606/c
 ID AA25606 standard; DNA; 1565 BP.
 XX
 AC AA25606;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DB Contig 005 from cosmid pKOS023-27 from *Streptomyces venezuelae*.
 XX
 KW Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 002;
 XX ketolide; PICCV; pteromycin; antibiotic production; narbomycin; de.
 OS *Streptomyces venezuelae*.
 XX
 FH Key Location/Qualifiers
 FT CDS 49..1507
 FT /*tag= A
 FT /product= "PICCV"
 XX
 PN W09961599-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 27-MAY-1999; 99WO-US011814.
 XX
 PR 28-MAY-1998; 98US-0087080P.
 PR 28-AUG-1998; 98US-00141908.
 PR 22-SEP-1998; 98US-0100880P.
 PR 08-FEB-1999; 99US-0119139P.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX
 PI Ashley G, Beltrach MC, Beltrach M, McDaniel R, Tang L;
 XX

DR WPI, 2000-072618/06.
DR P-PSDB; AAY67210.

PT New recombinant DNA encoding a domain of narbonolide polyketide synthase,
PT for production of ketolide antibiotics.

XX Disclosure; Page 38-39; 98pp; English.

XX This is config 005 from the recombinant cosmid pKOS023-27 DNA sequence
CC (see AAZ66001) which contains a Streptomyces venezuelae DNA insert. The
CC cosmid contains open reading frames which encode the various modules of
CC the narbonolide polyketide synthase (PKS). The invention relates to
CC recombinant DNA containing a coding sequence for a narbonolide PKS.
CC Polyketides are compounds synthesised from 2-carbon units through a
CC series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picrotoxin.
CC The narbonolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonolide PKS
CC (PICAI, PICAI1, PICAI2 and PICAI3). PICAI includes the loading module
CC and extender modules 1 and 2, PICAI1 includes extender modules 3 and 4,
CC PICAI2 includes extender module 5 and PICAI3 includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase domain
CC is found on the PICB protein. The nucleotide sequences encoding all of
CC these proteins can be isolated in recombinant form from the recombinant
CC cosmid pKOS023-27. Narbonolide is desosaminylated in *S. venezuelae* to
CC yield narbomycin, the desosaminyl transferase enzyme is required for this
CC conversion, and the desosaminyl transferase genes are also found in
CC cosmid pKOS023-27. The recombinant DNA of the invention is used to
CC express, in transformed cells, narbonolide (or its derivatives) or other
CC ketolides (particularly hybrids), which may then be converted (e.g. by
CC other enzymes recombinantly expressed in the same hosts) to polyketide
CC antibiotics or their intermediates. The antibiotics are useful in human
CC or veterinary medicine

XX Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;

Query Match 8.2%; Score 729.4; DB 3; Length 1565;
Best Local Similarity 68.1%; Pred. No. 7.5e-91;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGCGGGTGCACGCGCTTCAAGCGCGGATCAGGCGCGCGATGATCCACCGCTCG 2428
DB 1530 GGGGGGATCGGGGCGGGTGGCGGTCAGCGCAGAGAGCGCGGGCTCTCCACGCGTCC 1471
QY 2429 GCGACGCTCCGCTTCAAGTGTGTTGAGCGCGCGGCTCAACGATGTGTGAACCGTGAAG 2488
DB 1470 GCGCGGCTCGGCTTCAAGTGTGTTGAGCGCGCGGCTCAACGATGTGTGAACCGTGAAG 1411
QY 2489 AAGAACTCGTCCCGCGCTCGGATCGATGCTGCGCGCGCGATGAGGAGTGTGAGCG 2548
DB 1410 AAGTACTCGTCCCGCTCGGATCGATGCTGCGCGCGCGATGAGGAGTGTGAGCG 1351
QY 2549 ACCGAGTGCAGGCTCCGCTCGGATCGGAGTGCAGCGCGCGCGATGAGGAGTGTGAGCG 2608
DB 1350 ACCTCCGATGAGGAGTGTGCGGGGTCAGCGCGCGCGCGATGAGGAGTGTGAGCG 1291
QY 2609 AGCGCGGAGAAACCGCGCTCCGCTCGGATGAGTGAAGTGTGCGCGCGATGAGTGTGAGCG 2668
DB 1290 AGGTCCGAGAAACCGCGCTCCGCTCGGATGAGTGAAGTGTGCGCGCGATGAGTGTGAGCG 1231
QY 2669 GCCACCTGCGGATGAGGAGTGTGCGCGCGATGAGTGTGCGCGCGATGAGTGTGAGCG 2728
DB 1230 GCCACCTGCGGATGAGGAGTGTGCGCGCGATGAGTGTGCGCGCGATGAGTGTGAGCG 1171
QY 2729 TCGACACCGCGAGCGAGCTTCCAAACGCTGAACCGAGTGTGCGCGCGATGAGTGTGAGCG 2788
DB 1170 TCGCGCCCGGATGAGGAGTGTGCGCGCGATGAGTGTGCGCGCGATGAGTGTGAGCG 1111
QY 2789 CGCTCCGCGCGATGAGGAGTGTGCGCGCGATGAGTGTGCGCGCGATGAGTGTGAGCG 2848
DB 1110 CGCTCCGCGCGATGAGGAGTGTGCGCGCGATGAGTGTGCGCGCGATGAGTGTGAGCG 1051
QY 2849 TCGACACCGCGAGCGAGCTTCCAAACGCTGAACCGAGTGTGCGCGCGATGAGTGTGAGCG 2908

DB 1050 TGGCGGAGCTTGCCTGCTCAACGCGCGCTGTAGTCTCGGAGATGTTGAGGAGTCCATC 991
QY 2909 GCGCGGATGCGGCTGCACTGTGATGTTGCGCGATGAATGATGACAGGTGTGAGGCGG 2968
DB 990 GTCCTGCTTGCCTGCGCGCTGTTGAGGTGCGCGATGAATGATGACAGGTGTGAGGCGG 931
QY 2969 TTGGCGCGCGCGCGGAGATGATGTTGAGGCGCGGAGCGCGGATGCGCGCTTTCG 3028
DB 930 GAGGACGCGCGCGGAGAGATGATGAGCGAGCGCGGATGATGAGCGAGTGTGAGGCGG 871
QY 3029 GCGCGGATGCGGAGAGTCCCGGAGTGTTCCTGCAACCGCTTGAAGCGCGCGCTTTCG 3088
DB 870 GCGCGGATGCGGAGAGTCCCGGAGTGTTCCTGCAACCGCTTGAAGCGCGCGCTTTCG 811
QY 3089 CCGGTGCGCTTGTGATCTGCTGTGTTGAGCGCGGATGAGGAGTGTGAGGAGTGTGAGCG 3148
DB 810 CCGGTGCGCTTGTGATCTGCTGTGTTGAGCGCGGATGAGGAGTGTGAGGAGTGTGAGCG 751
QY 3149 AGGTCCGAGCGCGCGCTGCGCGCGAGCGCTGCTTTCGCGTGAAGGCGGAGTGTGAGCG 3208
DB 750 AGGCGCGAGCGCGCGCTGCGCGCGAGCGCTGCTTTCGCGTGAAGGCGGAGTGTGAGCG 691
QY 3209 TAGACGCTCAGGCGGAGCGCGCGCGCGCGCTGCGCGAGCGCGCGATGAGCGCGG 3268
DB 690 TAGACGCTCAGGCGGAGCGCGCGCGCGCGCGCTGCGCGAGCGCGCGATGAGCGCGG 631
QY 3269 TTGCTCAACGCTTTCAGGCGCGCGCGGAGGATGATGCGCGCTTTCAGCGCGCG 3328
DB 630 TTGCTCAACGCTTTCAGGCGCGCGCGGAGGATGATGCGCGCTTTCAGCGCGCG 571
QY 3329 TCGTTCATGACCGAGGATGAGATCTGTTGCGCGGAGCGCGAGCTTTCAGCGCGCG 3388
DB 570 TCGTTCATGACCGAGGATGAGATCTGTTGCGCGGAGCGCGAGCTTTCAGCGCGCG 511
QY 3389 GCACTGTGCAACCGCTTTCAGGCGCGGAGGATGAGCGCGGAGGATGAGCGCGG 3448
DB 510 GCGCGGATGCAACGCGGAGGATGAGCGCGGAGGATGAGCGCGGAGGATGAGCGCGG 451
QY 3449 CGGAGCTTTCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3508
DB 450 CGGAGCTTTCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
QY 3509 TTGAGGCGCTTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3568
DB 390 TTGAGGCGCTTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
QY 3569 CCGAGCTGCGGAGTGCATGAGGAGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3628
DB 330 CCGAGCTGCGGAGTGCATGAGGAGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 271
QY 3629 TCGAACCGAGCGCGCTTTCAGGCGCGGAGGATGAGCGCGCGCGCGCGCGCGCGCG 3688
DB 270 CCGAGCTGCGGAGTGCATGAGGAGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 211
QY 3689 AGTTCGCGAGCG 3748
DB 210 AGTTCGCGAGCG 151
QY 3749 GTGACCTGCGGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3808
DB 150 GCGAGGAGTGTGCG 91
QY 3809 TCGCGATTCGCTGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3859
DB 90 CGTTCG 40

RESULT 13
ADA09423/c
ID ADA09423 standard; DNA; 1565 BP.
XX
AC ADA09423;

XX 06-NOV-2003 (first entry)
XX S. venezuelae DNA encoding PICCV.
XX Streptomyces venezuelae; Sv; narbonolide polyketide synthase; PKS;
XX narbonolide PKS; narbomycin biosynthetase; picromycin biosynthetase;
XX PKS gene cluster; picromycin; glycosylation; hydroxylation;
XX C12 hydroxylase; PICK; desosamine biosynthetase;
XX desosaminyl transferase enzyme; antibiotic; narbonolide synthase; PICCV,
XX gene, ds.
XX Streptomyces venezuelae.
XX US6509455-B1.
XX 21-JAN-2003.
XX 07-SEP-2000; 2000US-00657440.
XX 30-APR-1997; 97US-00846247.
XX 06-MAY-1998; 98US-00073538.
XX 28-MAY-1998; 98US-0087080P.
XX 28-AUG-1998; 98US-00141908.
XX 22-SEP-1998; 98US-0100880P.
XX 08-FEB-1999; 99US-013139P.
XX 20-MAY-1999; 99US-0134990P.
XX 27-MAY-1999; 99US-00320878.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX WPI; 2003-352291/33.
XX Novel recombinant DNA compounds comprising coding sequences for
XX desosamine transferase gene of Streptomyces venezuelae, useful for
XX producing desosamine transferase which transfers desosamine to substrate
XX polyketides.
XX Discloure; Col 75-78; 132pp; English.
XX The present invention relates to recombinant DNA compounds that encode
XX Streptomyces venezuelae (Sv) narbonolide polyketide synthases (PKSs). The
XX recombinant PKSs are derived from narbonolide PKS and other genes
XX involved in narbomycin and picromycin biosynthesis in recombinant host
XX cells. The invention also discloses the S. venezuelae PKS gene cluster
XX that results in the production of picromycin. Also disclosed are enzymes
XX such as those responsible for glycosylation and hydroxylation, (e.g. C12
XX hydroxylase (PICK)), desosamine biosynthetase, and desosaminyl transferase
XX enzymes. The recombinant narbonolide, narbonolide derivatives, and
XX polyketides are useful as antibiotics and as intermediates in the
XX synthesis of compounds for pharmaceutical applications. The present
XX sequence encodes S. venezuelae PICCV.
SQ Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;
Query Match 8.2%; Score 729.4; DB 8; Length 1565;
Best Local Similarity 68.1%; Pred. No. 7.5e-91;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;
QY 2369 GGGGCGGTGACGCGGTCAGACGGCGGAGTCAGAGCCCGATGCCACCGCGT 2428
DB 1530 GGGGGAAATCGGGGCGGGTGGGTGACGCGAGAGGCGGGGCTCTCCACCGCGTCC 1471
QY 2429 GCGACGTCCCGTTCCAGTTGTTGAGCGGGGCGGTACCGAGTCGATCGAAACCGTCGAGG 2482
DB 1470 GGGGCGTGGCGCTCCAGCTGTGTTCAAGCGGGGCGGTGAGCACTGATCGAAGCGGTCAATG 1411
QY 2489 AAGAACTGTCTCCCGCGCTGCGAATGATCTGCGGCGCGACGTGACGAAAGTCGTGAGCG 2548
DB 1410 AAGTACTGTGCGCGCTCGACGGCGCCGCACTGCGCGCGCGCTCGACGAAAGTCCTCGTACG 1351

QY 2549 ACCGATGTCAGAGCTCCGCTCCGGGTGACCCGGCGCGACGTAGCGGGTGGCCCCCGCC 2608
DB 1350 ACCTCGGGTGAGGAGAGTGTGAGGGGTCAACCGGCGCCGATGATGAGGGGTTCGCGCTCC 1291
QY 2609 AGCCCGGGGAAACCGGCTCTCCCGGTACAGGTAAACGCTCCGAGCAGGTGATGTCAGC 2668
DB 1290 AGGTCCGGGAAAGCGGCTTCGCGGTACAGGTACGTCGCGAGAGATGACCTTGACAC 1231
QY 2669 GCCACTGCGAATGAGCGGTGCGGCGCATGTTTCCGGCGGATCCGCAACAGCTGGAGCG 2728
DB 1230 GCGACCTGCGGGTCCGGGTGCGGCGCATGTTGCGGCGGCTTGAATCCGACAGTTCCGGCG 1171
QY 2729 TCAGACCCCGACGACGCTCTCCAGCGCTAACCGAGTCGATGCTGATGCTCCGGGCTC 2788
DB 1170 TCGGCGCCCGGTGCGCAGGCTGTTCAAGGCGGTAGCGTATGATGATGAGTCCGAGGAGTG 1111
QY 2789 CGCTCCGGCGCGTATGATGAGAAACCGGGGCGAGGCGCTCCGCAACCTGATCCGTCGCC 2848
DB 1110 CGCTCCGGGAAACCGGCTCTGAAAGCGGTGAGGCGCTCGAGAGCTCGGCGCTCTCC 1051
QY 2849 TCAGAACACTGCGCGTCCGCGACCGCTGTAGTCTCGCGGACGCTGACGAAAGTCGAGC 2908
DB 1050 TCGGCGAGCTTTCGCTGTACAGGCGCGCTGTAGTCTCGGAAAGTTGACGAAAGTCGATC 991
QY 2909 GCGCGGTGCGGACCTGACCTGTTGAGTTGCGCGATGAAATCGACAGGTGATGAGCGCG 2968
DB 990 GTCTGCGCTGCGCGCGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 931
QY 2969 TTGGCGCGCGCGCGGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028
DB 930 GAGGCAAGCGCGCGGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 871
QY 3029 GCGGCGATCCGACGAAAGTCCCGGAGGTTCTTCTGACCGGTTGAAAGCGCCACGCTTG 3088
DB 870 GCGCGAGCTGCTGAAAGCGCGCGAGTCTCGGAGACGCGCGGAGCGCGCTTCTTG 811
QY 3089 CCGGTGATGCGCTGTGATCTGCTGTTTACCGCTGACAGGAGTACGAGACCGCGCG 3148
DB 810 CCGGTGATGCTGATCTGCTGCTGTTGAGCGGTGAGCGGATGAGCGGATGAGCGGCTG 751
QY 3149 AGGTCCACAGGCGCGGCTGCGCGCGCGCGCGCGCTGTTGATGAGGCGGAGCGGTTG 3208
DB 750 AGGCGCCAGAGCGCGGCTGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCG 691
QY 3209 TAGACGTGACGAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268
DB 690 TAGACGTGAGCG 631
QY 3269 TTGATCAAGGTTCCAGGCGCGCGGAGAGTACATCGCGGTGCTCGACGAGCAC 3328
DB 630 TTGATGAGCGGCTCAAGCGCGCGCGGAGAGTACATCGCGGAGGTTGCCCGCGGTATC 571
QY 3329 TCGTGTATGACCGAGGTGAGATCTCGTTGCGGCGACCAACGACTCCGCGTCTGATCT 3388
DB 570 TCGTGTATGACCGAGCGCGGACATGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCG 511
QY 3389 GCACCTGTCACCGGTCAGAGTACAGAGTACAGGAGGAGCATGTCAGTCCGGCGCGGATAGC 3448
DB 510 GCGCGGTGACACGAGCGAGATGAGCGAGGAGGAGCATGTCAGTCCGGCGCGGATAGC 451
QY 3449 CCGACGTGTACCGGAGAGCGGCTCCCGCGGTACCGGAGCTTCACAGCGCGCGCGCG 3508
DB 450 CCGACGTGTACCGGAGAGCGGCTTCGCGAGCGCGCGCGGTGAGAACCGCGCGCTGT 391
QY 3509 TTACGCGGCTGATGATGTTGCTCCAGTACTTTCGCGCGCGCGCGCGCGCGCGCG 3568
DB 390 TCGAGCGGAGACAGGCTGTTCTTCAGATACCGCGCGCGCGCGCGCGCGCGCGTTC 331
QY 3569 CCGAGCTGCGGAGTACGATGAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 3628
DB 330 CCGAGCTGCGGAGTACCGTCCAGAACAGGCGAGAGGCGCGCGAGAGCGCTCCGCGTCA 271
QY 3629 TCAGAACGAGCGCGCTGCTCCAGCGGAGTGAACGCGAGAGCGCGCTGATGAGCGGTG 3688

Db 270 CCCAGGTCTGTCGGGGCTCTCCAGCGGGGTGAAGGGCTGTTCCTGAGCCGACCGGCG 211
Qy 3689 AGTTGGCGGAGCCCGGACCGCTCTGCTTCGGCGGTTCGATGCGGACCGGTGAGCGG 3748
Db 210 AGCCGAGCGAGTGGCGGGCGGTGCTTCGGCTCTGTCGGCGGACGAGGCCCGCGCG 151
Qy 3749 GTGACCTCGGCGCTGAGCGCCGACGCGCGCGGTTCGGGTCTGTCGGCGG 3808
Db 150 GCGAGGGTCTGCGCGGACCGGCTGACCGCGCGCCCGACGATCGCTCCGGGGTCCGCGAG 91
Qy 3809 TCCGCGATCTTCGTCGCTACCGGCGGTTCGCGGTGAGTGTCTTCATCGAC 3859
Db 90 CGTTCCGCGCGGGCGGTGCGGAGAGGGCGGGCGGTCAATCGGAGCGTC 40

RESULT 14
ADH53467/c
ID ADH53467 standard; DNA; 1565 BP.
XX
AC ADH53467;
XX
DT 25-MAR-2004 (first entry)
XX
DE 5. venezuelae PK08023-26 cosmid DNA contig 005.
XX
KM Narbonolide polyketide synthase; PKS; desosamine;
KM desosaminyl transferase; beta-glucosidase; pick hydrolase; antibiotic;
KM pharmaceutical; gene therapy; antibacterial; infection; gene; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 50..1507
FT /*tag= a
FT /product= "PICCV protein"
XX
PN US2003162262-A1.
XX
PD 28-AUG-2003.
XX
PF 29-MAY-2002; 2002US-00160539.
XX
PR 28-MAY-1998; 98US-0087080P.
PR 28-AUG-1998; 98US-00141908.
PR 22-SEP-1998; 98US-0100880P.
PR 08-FEB-1999; 99US-0119139P.
PR 20-MAY-1999; 99US-0134990P.
PR 27-MAY-1999; 99US-00320878.
PR 07-SEP-2000; 2000US-00657440.
XX
PA (ASHL/) ASHLEY G.
PA (BETL/) BETLACH M C.
PA (BETL/) BETLACH M.
PA (MCDA/) MCDANIEL R.
PA (TANG/) TANG L.
XX
PI Ashley G, Betlach MC, Betlach M, Mcdaniel R, Tang L;
PI
DR MPI; 2003-897933/82.
DR P-PDB; ADH53453.
XX
PT New recombinant DNA compounds encoding a narbonolide polyketide synthase,
PT useful for expressing recombinant polyketide synthase genes in host cells
PT for the production of narbonolide and polyketides useful as antibiotics.
XX
PS Disclosure; SEQ ID NO 24; 136pp; English.
XX
CC The present invention relates to novel recombinant DNA compound that
CC encodes a domain of a narbonolide polyketide synthase (PKS), a desosamine
CC biosynthetic gene, a desosaminyl transferase gene or a beta-glucosidase
CC gene of Streptomyces venezuelae, or a pick hydrolase gene of S.
CC venezuelae. The recombinant DNA compounds are useful in expressing

CC recombinant polyketide synthase genes in host cells for the production of
CC narbonolide or its derivatives and polyketides that are useful as
CC antibiotics and as intermediates in the synthesis of compounds with
CC pharmaceutical value. They may be used for treating e.g. bacterial
CC infections. The invention is also useful in gene therapy. The present
CC sequence is Streptomyces venezuelae PK08023-26 cosmid DNA 005 contig used
CC in the invention.
XX
SQ Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;
Query Match 8.2%; Score 729.4; DB 10; Length 1565;
Best Local Similarity 68.1%; Pred. No. 7.5e-91;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;
Qy 2369 GGGCGGATCGAACCGCTTCAGAGCGCGCGGATCAGCGCCCGATGCTATCCACCGTCG 2428
Db 1530 GGGGGGATCGGGCGGGGTGCGGGTCAAGCGAGAAAGCGGGCGCTCTCCAGCCGCTCC 1471
Qy 2429 GCGACGTCGGCTTCGAGTGTGAGCGCGGGCGGTCACGACTGTGTGAAACCGTCGAGG 2488
Db 1470 GCGGCTTCGCGCTTCAGCTGTTCAGCGGGCGGTGAGACTGTATGAAAGCGTCCATG 1411
Qy 2489 AAGAATCTGTCCTCCGCGCTGCGATGATGCTGCGCGCCGACGTGACGAAGTCGTGAGC 2548
Db 1410 AAGTACTGTCGCGCGTCAAGCGCGCACTCGCGCGCGCTGCAAGAGTCCCTGACG 1351
Qy 2549 ACCGAGTGAAGCTCCGCTCCGGGTGACCCGCGCGCGACGTGAGCGGTGCGCCCGCC 2608
Db 1350 ACCTCGGTGAGGAGGTGTGCGGGGTCAACCGCGCGCGATGTATGCGGGTCCGCGCTCC 1291
Qy 2609 AGCCGGGGGAAACCGCGCTCCCGGTACAGTAAACGTCCTCCGAGAGGTGATCTGAC 2668
Db 1290 AGGTGGGAGAGCGGCTCGCGGTACAGTAAACGTCCTCCGAGAGATGACCTTGAC 1231
Qy 2669 GCCACCTGCGATGAGCGGTGCGGCGCATGCTTCGCGCGGATCCGCAACAGCTGAGCG 2728
Db 1230 GCGACCTGCGGAGTGGCGGGTGGCGCCGATGTGGCGGGCTTATCCGACAGTTCGCGG 1171
Qy 2729 TCGACACCCCGACGACAGCTCTTCAACCGGTAAACGATGCTGTCGATGCGCGGGTC 2788
Db 1170 TCGGCCCGGTGCGGAGCTGTTCAGGGCGTACCGGTATGATGTATGATGATCGGGGGTG 1111
Qy 2789 CGCTCGCGCGCGTGAAGTCAAGAACCGGGCGAGCGCTCCCGCAACTCGCTCGCTCC 2848
Db 1110 CGCTCGCGACCGCGCTCTTCAAGAGCGGTGAAGGCGCTCTGAGCTGCGCGCTCTCC 1051
Qy 2849 TCGCAACACTGCGCTGCTCCGACCGCTGTATCTTCGCGGACCGGTGACGAAGTCGAGC 2908
Db 1050 TCGGGAGCTGCGCGCTGCTCAAGCGCGCTGTATCTTCGGAATGTGACGAAGTCGATC 991
Qy 2909 GCGCGGTGCGGCGTGAAGTCTGTTGAGTTCGCGGATGAAGTCAACAGTTCGATGAGCGG 2968
Db 990 GTCTCTCCCTGCGCGCGCTGCTGTTGAGTTCGCGGATGAAGTCAACAGTTCGACAGCGG 931
Qy 2969 TTGGCCCCCGCGCGAGATGATGATGTGTTGAGGCGGACCGGACCGGTGCGCGCTTCG 3028
Db 930 GAGGCAAGCGCGCGGAGCAAGATGAGCGGAAGCCGAGGTGATCGCGCATCGCGCTCG 871
Qy 3029 GCGCGCATCGCGACGAGATCCGGAAGTCTTCTCTGACCGGTTCGAAAGGCGCACGCTTG 3088
Db 870 GCGCGAGCTGCTGGAAGCGCGCGAGTTCGCGGACCGCGGGAAGGCGGCTTCTTG 811
Qy 3089 CCGGTGTGCGCTGTACTGCTGCTGTTTCAAGCCGTCACGAGGATGAGGCGGATGCGG 3148
Db 810 CCGGTGTGCTGCTGTAATCTCTGCTGTTGAGGCGGTGAGCGAGGTGCGGATGCGGTGC 751
Qy 3149 AGGTCCCAAGCGCGCGCGTGGGCGGACCGTCTTCTCGTACGAGGATGAGGCGGATGCGG 3208
Db 750 AGGCCCCCAAGCGCGCGCGTGGGCGCTTCAGAGTGTGCGCTGATGAGGAGGATTCGTG 691
Qy 3209 TAGACGATGAGGCGGAGCGCGCGCGGCGGCGGTGAGACCAACCGCGATACCGCGG 3268
Db 690 TAGACGATGAGGCGGAGCGCGCGGTGCTGCTGCGGTGCGCGGCGGACGCTCCCGAGGCGCGG 631

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:24:14 ; Search time 43615.7 Seconds
(without alignments)
11760.789 Million cell updates/sec

Title: US-10-611-442-1
Sequence: 1 gcgcgccttcgacccatcgcg.....gtcgagcagcggagagatcc 9024

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBml:*

- 1: gb ba:*
- 2: gb in:*
- 3: gb env:*
- 4: gb om:*
- 5: gb ov:*
- 6: gb pac:*
- 7: gb ph:*
- 8: gb pr:*
- 9: gb ro:*
- 10: gb sta:*
- 11: gb sy:*
- 12: gb un:*
- 13: gb vi:*
- 14: gb hcg:*
- 15: gb pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
C 1	3037.4	34.0	8160	1	SEERYGENE	Y11199.1	S. erythraea	
C 2	3037.4	34.0	8160	6	BD073332	BD073332	Gene for	
C 3	3037.4	34.0	8160	6	AX000487	AX000487	Sequence	
C 4	3032.6	33.9	8051	6	AR092722	AR092722	Sequence	
C 5	3032.6	33.9	8340	1	SEU7459	SEU7459	Saccharopol	
C 6	2606.8	29.2	47981	1	AF63245	AF63245	Micromono	
C 7	2606.8	29.2	47981	6	AX112026	AX112026	Sequence	
C 8	1922.2	21.5	61845	1	AY623658	AY623658	Aeromicro	
C 9	1513.6	16.9	210614	1	AB088224	AB088224	Streptomy	
C 10	1313.8	14.7	17665	1	AF521878	AF521878	Streptomy	
C 11	1309.4	14.6	12441	1	AF079762	AF079762	Streptomy	
C 12	1309.4	14.6	12441	6	BD217172	BD217172	DNA encod	
C 13	1309	14.6	58343	1	AB089954	AB089954	Micromono	
C 14	1276	14.3	9523	1	AF237894S1	AF237894	Streptomy	
C 15	1269.8	14.2	15052	1	AF055579	AF055579	Streptomy	
C 16	1241.2	13.9	60196	6	AX697977	AX697977	Sequence	
C 17	913.8	10.2	1206	6	BD073333	BD073333	Gene for	
C 18	913.8	10.2	1206	6	AX000494	AX000494	Sequence	

19	884	9.9	8407	1	SEU82823	US82823	Saccharopol
20	729.4	8.2	1565	6	BD232539	BD232539	Recombina
21	729.4	8.2	1565	6	AR271685	AR271685	Sequence
22	729.4	8.2	1565	6	AR277669	AR277669	Sequence
23	729.4	8.2	1565	6	AR679279	AR679279	Sequence
24	722	8.1	1458	6	BD217175	BD217175	DNA encod
25	720.6	8.1	85915	1	AY509120	AY509120	Streptomy
26	712.2	8.0	3439	1	SEV14332	Y14332	Saccharopol
27	712.2	8.0	3439	6	BD073330	BD073330	Gene for
28	712.2	8.0	3439	6	AX000482	AX000482	Sequence
29	703	7.9	3459	1	SEU7454	U77454	Saccharopol
30	701.4	7.8	3756	6	AR092721	AR092721	Sequence
31	673.2	7.5	61845	1	AY623658	AY623658	Aeromicro
32	667.2	7.5	1467	6	AX698013	AX698013	Sequence
33	596.6	6.7	1227	6	AX698011	AX698011	Sequence
34	595.4	6.7	1248	6	BD217174	BD217174	DNA encod
35	564.8	6.3	109528	1	AF040570	AF040570	Amycolato
36	547	6.1	9523	1	AF237894S1	AF237894	Streptomy
37	545.6	6.1	39428	1	AF322256	AF322256	Streptomy
38	545.4	6.1	15052	1	AF055579	AF055579	Streptomy
39	540.8	6.0	34869	1	AF324838	AF324838	Streptomy
40	534.4	6.0	18977	1	SAR7932	AJ007932	Streptomy
41	519.2	5.8	5970	6	BD232536	BD232536	Recombina
42	519.2	5.8	5970	6	AR271682	AR271682	Sequence
43	519.2	5.8	5970	6	AR277666	AR277666	Sequence
44	519.2	5.8	5970	6	AR679276	AR679276	Sequence
45	500.4	5.6	45055	6	AX574197	AX574197	Sequence

ALIGNMENTS

RESULT 1	SEERYGENE	8160 bp	DNA	linear	BCT 18-APR-2005
LOCUS	S. erythraea erythromycin gene cluster DNA.				
DEFINITION	Y11199				
ACCESSION	Y11199.1	GI:2665334			
VERSION	2,3 dehydratase; aminotransferase; dehydratase; deoxyhexose				
KEYWORDS	reductase; eryBIV gene; eryBV gene; eryBVI gene; eryBVII gene;				
	eryCIV gene; eryCV gene; eryCVI gene; erythromycin;				
	glycosyltransferase; TDP-4-keto-6-deoxyglucose 3,5 epimerase;				
	TDP-4-keto-6-deoxyglucose 5 epimerase;				
	TDP-N-dimethyldeoxosamine-N-methyltransferase.				
SOURCE	Saccharopolyspora erythraea				
ORGANISM	Bacteria; Actinobacteridae; Actinobacteriales; Saccharopolyspora				
REFERENCE	1	Gaiser, S., Bohm, G.A., Cortes, J. and Leadlay, P.F.			
AUTHORS	1	Analysis of seven genes from the eryA-eryK region of the			
TITLE	erythromycin biosynthetic gene cluster in Saccharopolyspora				
JOURNAL	erythraea				
PUBMED	Mol. Gen. Genet. 256 (3), 239-251 (1997)				
REFERENCE	2	(bases 1 to 8160)			
AUTHORS	Leadlay, P.F.				
TITLE	Direct Submision				
JOURNAL	Submitted (12-FEB-1997) P.F. Leadlay, Department Of Biochemistry,				
COMMENT	University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK				
FEATURES	Related sequence: J05776.				
source	Location/Qualifiers				
	1..8160				
	/organism="Saccharopolyspora erythraea"				
	/mol_type="genomic DNA"				
	/strain="NRRL2338"				
	/db_xref="taxon:1836"				
	242..1210				
	/gene="EryBIV"				
	242..1210				
	/gene="EryBIV"				
	/function="Involved in mycrose biosynthesis"				
	/codon_start=1				

[illegible][illegible]

Db 7607 GTGGGGGTGAACCTGAAGAACCCGCCGACCGCCAGCTCACTGACCACTCCGGTTC 7548
QY 2388 AGACGGGCGGGATCAAGGCCCCGAAAGTATCCACCCCTGGGGAGGTCCGGTTGAGTT 2447
Db 7547 ACCTTCGGCCGAGGAAGCCCGGGTGTCTCTCCACCCCTGGCGATCTCTGTCGAGCT 7488
QY 2448 GATTGAGCCGGGCGGTCAACGACTGTGTGAAACCGTTCGAGAAAGAACTCTCCCGGCT 2507
Db 7487 GATTGAGCCGTGGCGGTCAACGACTGTGTGAAACCGTTCGAGAAAGAACTCTCTCCCGGGGC 7428
QY 2508 GCGGATCGATGCTGCGCCCGGAGTGAAGAGTCTGAGACGACCGAGTCAAGGCTCCGGT 2567
Db 7427 GGGGCTGCAAGCCCTCGTTCTTCCAGCAAGAGTCTCCACACTCCGCGAGGCTGATCG 7368
QY 2568 CCGGGGTGACCCGCGCGGCGACGATGAGGGGTGGCCCCCGCAGCCCGGGGAAACCGGCT 2627
Db 7367 ACCGGGTGACCCGCGCGGCGGATGAGGGGTGGGCCCTCCAGCTCCGGGAAAGCCGCT 7308
QY 2628 CCGGATCAAGATGAGCGTCTCCGAGCAGTCTGATCTGCAACCGCCTCTGCGGATGGGCG 2687
Db 7307 CCGGATGAGATGAGCGTCTCCGAGCAGTCTGATCTGCAACCGCCTCTGCGGATGGGCG 7248
QY 2688 TCGGGCGCATCGTTTCGCGCGGATCCGAAACAGTGGGCGTGCACCCCGACGAGGC 2747
Db 7247 TGGAGCGCATCGTCTCCGCGCGGATGGCAGAGTCTGCGCTCCACCCCGCGAGGC 7188
QY 2748 TCTCCAAACGCGTAAACCAAGTCCGCTCTGATGATCCCGGGGTCCCTCGCGCGGCTGATCGA 2807
Db 7187 TCTCCAAAGGCGTAAACCAAGTCCGATGATGATCCCGGGGTCCCTCGCGCGGCTGATCGA 7128
QY 2808 CGAAGCGGCGAGGACCTCCCGCAACTCGGTCCGTTCCGCTCGGACAACTTCGCGCT 2867
Db 7127 CGAAGCGGCGAAGCCCTCCCGCAAGTCTGAGGCTCTGAGATCGAAGCCGCGCTCTCT 7068
QY 2868 CCGGACCGCTGATGCTCTCGCGGACGATGACGAAGTGAAGCGGCGGCTGGGCTGACT 2927
Db 7067 CCGGCGCGCTGATGCTCTCGCGGACGATGACGAAGTGAAGCGGCGGCTGGGCTGACT 7008
QY 2928 CGTTGAGTTCGCGGATGAAAGTTCGACAGTTCGATGAGGCGGTTGGCCCGCGGACAGA 2987
Db 7007 CGTTGAGTTCGCGGATGAAAGTTCGACAGTTCGATGAGGCGGTTGGCCCGCGGACAGA 6948
QY 2988 TGAATGATGAGGCGGAGCGGACCGGATCCGCGCTTCGCGCGGATCCGAGGAAT 3047
Db 6947 TGAATGATGAGGCGGAGCGGATCCGCGCTTCGCGCGGATCCGAGGAAT 6888
QY 3048 CCGGAGGATCTTCTCTGACCGGTTGAAAGCGCACGCTTTCGCGGATGAGCTGATCT 3107
Db 6887 CCGGAGGATCTTCTCTGACCGGTTGAAAGCGCACGCTTTCGCGGATGAGCTGATCT 6828
QY 3108 CGTCTGTTTCAAGCCCGTACAGGAGGATGACCGCGCGGAGGTTCCACAGGCGCGGCT 3167
Db 6827 CGTCTGTTTCAAGCCCGTACAGGAGGATGACCGCGCGGAGGTTCCACAGGCGCGGCT 6768
QY 3168 GGGGCGGAGCGCTCTGTTGAGTGAAGGCGGAGGCTTGTGATGAGCGGTCAAGCG 3227
Db 6767 GGGGCGGAGCGCTCTGTTGAGTGAAGGCGGAGGCTTGTGATGAGCGGTCAAGCG 6708
QY 3228 CCGGCGGAGCGGAGTGAAGGCGGAGGCTTGTGATGAGCGGTCAAGCG 3287
Db 6707 CCGGCGGAGCGGAGTGAAGGCGGAGGCTTGTGATGAGCGGTCAAGCG 6648
QY 3288 CCGCGGAGGATGATGATGCGCGGTGGGTTCTCCGACGCGCACCTGTGATGATGCGAGTGA 3347
Db 6647 CCGCGGAGGATGATGATGCGCGGTGGGTTCTCCGACGCGCACCTGTGATGATGCGAGTGA 6588
QY 3348 GGAATCTGTTGCGGCGGACCAACGATCCGCGGTGATCTTGGACCTGATCCCGTACCG 3407
Db 6587 GCGTCTCTGTTGCGGCGGCGGAGCGGATGCGGCTCTGATGAGCGGACCGGTCACCGGACG 6528
QY 3408 AGAAGTGAAGCGGAAATGACAGTGGGCGGAGGATGAGCGCGGACGCTGATCGGAAAG 3467
Db 6527 AGAAGTGAAGCGGAAATGACAGTGGGCGGAGGATGAGCGCGGACGCTGATGAGGAAAG 6468

QY 3468 CCGGCTCCGCGCGTACCGAGGCTCCAAAGCGCGCGGCGTTCAAGCGGCTGATCGGT 3527
Db 6467 CAGGCTTGCAGTGAAGCGCGGCTCCAGTGGCGCTTGGGCGTCCAGCGGCTTGAATCGGT 6408
QY 3528 TGTCTCAAGTACTTGGCCCGCGGCGGCTGCTCAACCGCGGAGTCCAGTGGGAGTGGAT 3587
Db 6407 TGAACCAAGTACTTGGCCCGCGGAGCGGTGCTCAACCGCGGAGTCCAGTGGGAGTGGG 6348
QY 3588 CGAACAGTTCGAGCAGCTGCGCGAACCCCGGCGGCTTCAGGTGAACCGACGCGCATCT 3647
Db 6347 CGAACAGTTCGAGCAGTGGGCGCAACTCCGCGGCTTCAGGCGGAGCGCGCACGCGCTCT 6288
QY 3648 GCTCCAAACGGGGGAAACGCGAGGCGCGGATGAGGCGGATTTGGGCGAGCGGACCG 3707
Db 6287 GCTCCAAACGGGCGGAAACGAGTGGCGCGCGGATGAGGCGGAGTGGGCGAGCGGCGG 6228
QY 3708 CCTGTGCTTCGCGGCTGTCGAGTGGGAGCAGCGGATGAGGCGGAGTCTCGGCGTGAAG 3767
Db 6227 CTTGCAAGTTCGCGGCTGTCGAGTGGGAGCAGCGGATGAGGCGGAGTCTCGGCGTGAAG 6168
QY 3768 CCGGACCAACGCGCGGCGGATGCGGCTGAGTGGGCGGCTCCGATCTCCGCTGAT 3827
Db 6167 CCGGACCAACGCGCGGCGGATGCGGCTGAGTGGGCGGCTCCGATCTCCGCTGAT 6108
QY 3828 CCGCGGTCGCGGATGAGTGGTCTTATGACGCTGGAACCTCTTGGCGCTGTGTGTC 3887
Db 6107 GGGCGGTCGCGGATGAGTGGTCTTATGACGCTGGAACCTCTTGGCGCTGTGTGTC 6048
QY 3888 GAGGATCAAGAACCGTTGGCTTCCGCTTGTCCACTCCGCTTGTATCAACGACCGCTG 3947
Db 6047 GAGGATCAAGAACCGTTGGCTTCCGCTTGTCCACTCCGCTTGTATCAACGCTG 5988
QY 3948 GTGTGGCGAGTGCAGTGAAGTGCACACCGCGGATGCTCACTGACACCGAGGG 4007
Db 5987 GTGTGGCGAGCGGATGATGTCGACACCGCGGATGCTCACTGACACCGAGGG 5928
QY 4008 CCGGTGGGAGGCGGAGCACCGGTTGGGAGGCACTCGGTCTGTCTCAAGCGGAGT 4067
Db 5927 CCGGTGGGAGGCGGAGCACCGGTTGGGAGGCTGTGTCTGTCTCAAGCGGAGT 5868
QY 4068 GGGCTCGGTCGAGGAGCGGATCTGTGTGGGAGGCGGAGGAGTGAAGGCTGTGAGCG 4127
Db 5867 GGGCTCGGTCGAGGAGCGGATCTGTGTGGGAGGCGGAGGAGTGAAGGCTGTGAGCG 5808
QY 4128 ACCTTCTCGCGCGGAGATGCGCTGACAGCGGTCACGATGCGATGCGGATGCGGCTG 4187
Db 5807 ACCTTCTCGCGCTCGAGAGATGCGCTGACAGCGGATGCGATGCGGATGCGGCTG 5748
QY 4188 TCCACCAAGATGATCACTGATGATGTTGCTCTCTGTCGCGGCGGAGCGAGTGCAG 4247
Db 5747 TCCACCAAGATGATCACTGATGATGTTGCTCTCTGTCGCGGAGTGCAG 5688
QY 4248 GTGAGCGCGGATGAGTGCAGGCTGCTGTGTGATGAGGCGGATGAGCCGCTGTGTC 4307
Db 5687 GATATGCGCGGAGCGTGCAGGAGCTGTGTGATGAGGCGGATGAGCCGCTGTGTC 5628
QY 4308 TCCCTGTGCTTCGCGGACCGCTGAGGAGGATGAGCCGATGAGCGCGCGGACTGCTC 4367
Db 5627 ACCCTGTGCTTCGCGGAGGCGTGCAGGAGGATGAGCCGATGAGCGCGGACTGCTC 5568
QY 4368 ATCTTTCGCTGTGTGCGGATCTCGGATGACCACTTGTCTGGGCGGATGCGGAGTTGTC 4427
Db 5567 ATCTTTCGCTGTGTGCGGATCTCGGATGACCACTTGTCTGGGCGGATGCGGAGTTGTC 5508
QY 4428 ATGAGCGGATCGTTTGGCGGAGGCGGATGAGGAGTGAACCAACCGCGCGGCTTGAG 4487
Db 5507 ATGAGCGGATCGGTTGGCGGAGGCGGATGAGGAGTGAACCAACCGCGCGGCTTGAG 5448
QY 4488 GCGGTGACCGCTTGGTGGCGTGAAGCTGAACCACTTCGCGGTCACCGGATCCGCCAC 4547
Db 5447 GAGGTGACCGGCTTGGTGGCGTGAAGCTGAACCACTTCGCGGTCGGAAGGCGGAGCC 5388

QY 4548 GAGCTCCACCCGTCGTGACAGCCAGAGGCGGTGCGGCTGAGAGAGAGCTTGACCTTGG 4607
DB GGCCGCCCGCCGCGCGGTGACAGCCAGCGGTGCGGCGGTGAGAGAGAGCTTGACCTTGG 5328
QY 4608 TGGTCGCGCGGATCTTTCGCGAGCGCTTCCAGAGTGTGTGGTCCGCCACAGGTGTAACG 4667
DB TGGTCGCGCGGATCTTTCGCGAGCGCTTCCAGAGTGTGTGGTCCGCCACAGGTGTAACG 5268
QY 4668 CCGAGCATTCGCGCGGTGTCGCGGAGTGCAGCGCGGAGAGTGTGTGGTCCGCCACAG 4727
DB CCGAGCATTCGCGCGGTGTCGCGGAGTGCAGCGCGGAGAGTGTGTGGTCCGCCACAG 5208
QY 5267 CCGAGTATTCGCGCGGTGTCGCGGAGTGCAGCGCGGAGAGTGTGTGGTCCGCCACAG 5208
DB CCGAGTATTCGCGCGGTGTCGCGGAGTGCAGCGCGGAGAGTGTGTGGTCCGCCACAG 5208
QY 4728 CCGGTCCGCGGATTCAGAGTGCAGAGCAACCGGTGTGATTCGAGCCAGCTTCCGCGTGC 4787
DB CCGGTCTTCGCGGATTCAGAGTGCAGAGCAACCGGTGTGATTCGAGCCAGCTTCCGCGTGC 5148
QY 4788 GCGGTGCGCGGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGC 4847
DB GCGGTGCGCGGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGC 5088
QY 5147 GCGGTGCGCGGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGC 5088
DB AGCACAGATTCAGAGGCGAGCGGTGCGAGGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 4907
DB AGCACAGATTCAGAGGCGAGCGGTGCGAGGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 5028
QY 4908 AGGTGCGGAGCCCGGCTCGAGCTCCCGACAGAGGCGCGCGTGTGTGAGCGAGTGTG 4967
DB AGGTGCGGAGCCCGGCTCGAGCTCCCGACAGAGGCGCGCGTGTGTGAGCGAGTGTG 4968
QY 4968 TTTGTCAGAGGCGCACTCCAGAGGCGCGAGCAACCGTGTGCGGTGCGGTGCGGTGCGGTG 5027
DB TTTGTCAGAGGCGCACTCCAGAGGCGCGAGCAACCGTGTGCGGTGCGGTGCGGTGCGGTG 4908
QY 5028 CCCACGTGCGAGGCGGTGCGAGAGCGCTCGGAGCGCGAGAGTGCAGAGTGCAGAGTGCAG 5087
DB CCCACGTGCGAGGCGGTGCGAGAGCGCTCGGAGCGCGAGAGTGCAGAGTGCAGAGTGCAG 4848
QY 4907 CCCACGTGCGAGGCGGTGCGAGAGCGCTCGGAGCGCGAGAGTGCAGAGTGCAGAGTGCAG 4848
DB AGCGGCTTCAGCGCGGTGCGCGGCTGTGTGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCG 5147
QY 4847 GCGGCTTCATTCGCGCGGTGCTGTGTGTGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCG 4788
DB GCGGCTTCATTCGCGCGGTGCTGTGTGTGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCG 4788
QY 5148 GCTGAGTGTGACGTGAGT 5207
DB GCTGAGTGTGACGTGAGT 4728
QY 4787 GCTGAGTGTGACGTGAGT 4728
DB ACCGCTGTGCGGCGGCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5267
QY 5208 ACCGCTGTGCGGCGGCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5267
DB ACCGCTGTGCGGCGGCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4668
QY 4727 ACCGCTGTGCGGCGGCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4668
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5327
QY 5268 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5327
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4608
QY 4667 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4608
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5328
QY 5328 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5327
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4607
QY 4607 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4607
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4548
QY 5388 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5447
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4488
QY 4547 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4488
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5507
QY 5448 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5507
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4428
QY 4487 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4428
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4368
QY 4427 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4368
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5567
QY 5508 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5567
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5627
QY 5568 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5627
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4308
QY 4367 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4308
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5687
QY 5628 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5687

DB 4307 TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4248
QY 5688 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5747
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4188
QY 4247 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4188
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5807
QY 5748 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5807
DB TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4128
QY 4187 TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4128
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5867
QY 5808 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5867
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4068
QY 4127 TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4068
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5927
QY 5868 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5927
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4008
QY 4067 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4008
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5987
QY 5928 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5987
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3948
QY 4007 GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3948
DB GAGCGAATCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6047
QY 5988 TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6047
DB TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3888
QY 3947 TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3888
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6107
QY 6048 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6107
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3828
QY 3887 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3828
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3768
QY 6108 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6167
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3768
QY 3827 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3768
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6227
QY 6168 TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6227
DB TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3708
QY 3767 TGGTCGCGGTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3708
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6287
QY 6228 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6287
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3648
QY 3707 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3648
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3588
QY 3647 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3588
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6407
QY 6348 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6407
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3528
QY 3587 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3528
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6467
QY 6408 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6467
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3468
QY 3527 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3468
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6527
QY 6468 TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6527
DB TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3408
QY 3467 TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3408
DB TGGCGTCGTGCGCGGATTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6587
QY 6588 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6587
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3348
QY 3407 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3348
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3393
QY 3347 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3393
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6647
QY 6647 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 6647
DB GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3292
QY 3292 GATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3292

RESULT 2

[illegible]

QY	2388	AGACGGCCGGATCAGGGCCCGAAGCTGATATCCACCCGTCGGACAGCTCCGTTGCAATT	244
Dp	7547	ACCTTCCGCGAGGAAGCCGCGGTGTCCTCCACCCTCCGACATGCTCTGTTGAACT	7488
QY	2448	GATTGAGCTCGGCGCGGTCACCGACTGGTGGAAACCGTCGAGGAAGAATCTGTCCTCCGACT	2507
Dp	7487	GATTGAGCCGTGCGGTCAACCGACTGGTTCGAAAGCCGTGAGGAAGAATCTGTCGCGGGGC	7428
QY	2508	GCGGATCGAATGCTCGGCGCGGACGTGACGAAGTGGTGGACGACGGAATGCAAGGCTCCGGT	256
Dp	7427	GGGGCTGACGCCCCCTGTTTCTCCAGCACGAAGTTCTCCACCACTTCGCGCAGGCTGGTCG	7366
QY	2568	CCGGGGGTGACCCGCGCCGCGCAGACTGATAGCGGTGAGCCCCCGCCAGCCCGGGGAAACCGGCTT	2627
Dp	7367	ACGGGGGTGACCCCGGCCCGCGATGTATAGCGGGTGGCGCCCTCCAGGCTCCGGGAAAGCCCGCT	7308
QY	2628	CCCGGTACAGGTATACGTCCTCCAGCAGGTGATCTGTGACCGCCACCTGCGCATATGAGCTGG	2687
Dp	7307	GCGGTGAGAGGTATACGTCGCGCAGCAGGTGCACTGTGACCGCCACTGGGGGGTGCCTGG	7248
QY	2688	TGCGGCGCATGTTTCCGCGCGGATTCGCAACAGTGGGCGTTCGACACCCCGACGCAAGCC	2747
Dp	7247	TGGGACCGATGTCCTCCGGCCGGAATGGCGAGCACTGGGCGTCCACACCCCGCGCAGGC	7188
QY	2748	TCTCCAAACGCGTAAACCAAGGTCCGATGCTGCAATGCCCGGAGTCCGTCGCGCGGATGTCGA	2807
Dp	7187	TCTCCAGGGCGTAAACCAAGGTGCAATGTCATGCCCCGGGGTTCGCTCGGCGCGATGTCGA	7128
QY	2808	CGAACCCGGGCGAGGCCCTCCCGCGAACCTCGGCGGTTGCCCCCGGACAACTGCGCGTCGT	2867
Dp	7127	CGAACCCGACAGGCCCTCCGCGACGTCGTGTCGCTCGAGTCCGACAGCCGCGCGTCGT	7068
QY	2868	CCCGACCGCTGTAGTCTCTCGCGGACCGGTGACGAAGTCAGCGCGCGGTCGCGGTGACCT	2927
Dp	7067	CGCGGCGCGTGTAGTCTCTCGCGACCGTCAAGAAGTCAAGCGGCGGTTGCGCGGCTGACCT	7008
QY	2928	CGTTGAGTTGCGCGCATGAAATCGAACCGTGCATGAGGCGGTTGGCCCCCGCGCAGGA	2987
Dp	7007	CGTTGACCTCGGCGCATAAAGTCGACGAGGTGCGTGAACCGGTTCGCCCCGTCGCGCAGGA	6948
QY	2988	TGATGTGTTGAGGCGCGAGCCGGAACCGGTGCGGCGGTTGCGGCGCGCATCCGCGAGGAAGT	3047
Dp	6947	TGATGTGTTGAGGCGCGAGCCGGAATCGGCGGTCCTCGCGCGCATCCGCGAGGAAGC	6888
QY	3048	CCCGAGAGTTCCTTCGTGACCCGTTGGAAGGCGCGCTTGGCGGATGAGGCTGTGATCT	3107
Dp	6887	CTTCGACAGTTCCTTCGTGACCGGTTGGAAGGCGCGCGCTTGGCGGATGAGTCTGTGATCT	6828
QY	3108	CGTCGTGTTCAAGCCCGGTAACAGGGAAGGTTAAGAACCGGCGGAGGTCCCAACAGCGCGGCT	3167
Dp	6827	CGTCGTGTTCAAGCCCGGTAAGGGAAGGTCGGAATCGGCGCCAGCTCCCAACAGCGCGGCT	6768
QY	3168	GGCAGCGCAGCGTCTGTTCCGTGAGGAGCGGAAGCGTTGTGTAGACGCTCAGGCGGAGAC	3227
Dp	6767	GGCAGTTCAAGGTCGTCTCGGTGAGGCGGAAGGCTTGTGTGTAGACGCTGAGTCTGAAAC	6708
QY	3228	CCCGGCGGGGCGGTGCGACCAACGCCCCGATATCCCGGAGTTGTCAACGCTTCCAGGC	3287
Dp	6707	CGCGCCCGGGCGGTGCGACCAACGACTCGCGGAGCCGAGGAGTGTGTGAGCGGCTCGAGGC	6648
QY	3288	CGCGCGAGAGGTACATATGCGCGGTGGGGTTCTCCAGCGGCACTGCTGCAATGACGGAATGA	3347
Dp	6647	CGCGCGACATGTACATGCTTTCGGGTTTTCCTGGGCACTGTGCAATGATTCGCGGCA	6588
QY	3348	GGATCTCTGTTGCCGGGACCAACGACTCGCGGTCTGTACTTGGACACTGTCAACCCGTGACG	3407
Dp	6587	GCGCTCTGTTGCCCGCGGAGACCGAATGCGGCTCTGTGCGGGGACACCGGTCAACCCGACGC	6528
QY	3408	AGAAAGTACAGCGGAAACATGACGTGCGGCGCGGGGTGAGACCCGACGCTGTACGAGGAAG	3467
Dp	6527	AGAAAGTACAGCGGAAACATGACGTGCGGCGCGGGGTGAGAGCCGACGCTTATGAGGAAG	6468

QY	3468	CGGGCTCCCGCGGTAACGGCAAGCTTCTCAACGCGCGGGGGGTTTCAGGGGCTGATGTTG	3527
Db	6467	CAGGCTTGGCGGTAAACCGCGCGCGTTCAGGTGGCTTGGGGGTCCAGGGGCTTATGTTG	6408
QY	3528	TGCTCCAGTATTGGCCCGCGGGGGGCGTCTCAACCGCGGTTGCCCAAGCTTGGGGAT	3587
Db	6407	TGAAACAGTACTTTCGCGCGGAACTCGTCTCAACCGGGTCCCAAGTCCGGATGGGC	6348
QY	3588	CGAACAGGTGAGCAGCTCGCGGAACCCGCGCCGGTCCAGGTCCAAACCGAGCGCACT	3547
Db	6347	CGAACAGGTGAGCAGGTTGGGCGAACTCCGCGCGTCCAGGCGAGCGCGCAAGCGTCT	6288
QY	3648	GCTCCAAACGGGGTGAACGGCGAAGCGCGTATGTGTGGGGGTGTGATTTCCGCGAAGCCGACCG	3707
Db	6287	GCTCCAGGGGGTGAACGGGTGCCCGCTGTATGTTCGGCGGAAGGTTCGCGAAGCGGGCGGG	6228
QY	3708	CCTGTGCTTGGGCGGGTGTTCGATGCGGGAACCGGTGAGGGCGGGTGAACCTCCGCGTGAAG	3767
Db	6227	CTTGCAAGTCCGCGCTCACTTCACCCGAGCCGGTGTGACGGCGGAAGACTTCGAGCTCAAGCG	6168
QY	3768	CCCGACCAACGAGCGCGCGCGGTTCGGCGTTCGGTCCGCGTCCGCGAATCTCCGTCCGTGA	3827
Db	6167	CCCGACGACACCGCGCGCGCGGTTCGAAGTTCGGGCGCGCGCGTCCGCGAACCCTTCCT	6108
QY	3828	CGGCGGTTCGGGTGAGAAATGTCTTTCATCCAGTGTCCAAACCTTCTTGCGGTCTGTGGTGC	3887
Db	6107	GGGCGGTTCGGGTGAGAAATGTCTTTCATCCGGAAGCGAACTCTTCGCACTTGTGTGGGCA	6048
QY	3888	GAGGATTCAGAAACGGTTTGCGTTTCCGCTTTCGCCACTCCGCGGTGATCAACGACGCGT	3947
Db	6047	GGTCTCAACGAACCGTTTCCGCGTCTCTGTGTCCATTGCGGTGATTCAGTCTGCGCGCTG	5988
QY	3948	GTGTGTGGAGTCCGATGATGCTCGGCAACCCGCGGAATGTCCTCACTGAACCCGAGGG	4007
Db	5987	GTGTGTGGAGTCCGATGATGTCGACACCCGCGGAATGTCCTGCTGATGAACCCGCGGG	5928
QY	4008	CCGGTCCGGAAGGCGAGCAACCGTTTCGGCGAGCACTCGGTCTGTGTTCAGTCCGACGGGT	4067
Db	5927	CCGGTGGGCAAGCGCAGACACCCGGTCCGAGAGCTGTTCGGGTCTCCAGGCTCCGACGGC	5868
QY	4068	GGCTTCGGTGAAGGGGATCACTGTGTGGAGAGCGGGGAGAAATGAGGCTGTGGCAGCG	4127
Db	5867	GGCTGGTTCGGTGAAGGCTGTGCACTGTGTGGACCCCGGGGAAAGTGAAGTTGTGCAACA	5808
QY	4128	ACCTTCTCCGCGCGAGGATGCGCTGCAAGCGGTCACGATGATGCGGGTGGCGGTGCG	4187
Db	5807	ACCTTCTCCGCTCCGAGGATGGCTGTGCAACTGTGTGCGGTGATGTCGGGTGGCGCGGAG	5748
QY	4188	TCCACCAAGATGATCAAGTACTGTGATGTTGCTCTCTGTGCGCGGAGAGCGAGTGCAGCG	4247
Db	5747	TCCACCAAGATGATCAAGTACTGTGATGTTGTTCTGCTCCCAAGATTCMAACGCGTGCAGCG	5688
QY	4248	GTGAAGCGCGGATGCTCGCGGAGCTCGCGGTGTGAAGACGGGTGTTGACCGGGTGTGCG	4307
Db	5687	GATATGCGCGGCAAGTCCGGAAGCTCGTCCGAGTGAACCGCGTGTGTGAAGCGGTTGTGCG	5628
QY	4308	TCCCTGTGCTCCGCGCAACCGGTGAGGAGGTGAAGCCCATGTGCGCGCGCGCACTCGTCTC	4367
Db	5627	ACCTGTGTTCCGCGGAAGCGGTGAGCGAGGGTGAAGCCCATGCGCGCGGCGCACTCGTCTC	5568
QY	4368	ATCTTGGCGTTGTGTCCGATCTTCGTTACCACTTGTTCGGGCGGATGCCGAATGTTGTGCG	4427
Db	5567	ATCTTGGCGTTGTGTCCGCAATTCGTTCAACAGCTTGTTCGGGTGCATCCCAAGTTGTGCG	5508
QY	4428	ATGGCGCGGATCCGTTTCGGCGCAGAGCGCGTGTGCGGTGAACCAACGCGCGCGCCCTCGAAG	4487
Db	5507	ATGGCGCGGATCCGGTTCGCGCAGAGCCCGGTGTGCGGTGAACGATGGCGCGCGCTTCGAAC	5448
QY	4488	GGGTGTACCGGCTTGGTGGCGTGAAGACTGAACACTTCGGGCTCAACCGGATCCGCGCAC	4547
Db	5447	GAGGTGAACGCGCTTCTGTGGCGTGAAGACTGAACACTTCGGGCTTCGAAGCGCGCGAC	5388
QY	4548	GGACGTCAACCGGTGTGAGCGCCAGAGCGTGGCGCGGTTCGAAGAAGACTTGACTTGA	4607

Db	5387	GGCGCCCGCCGCGCGGTGACGCCACGCGGTGCGCGGCTCCAGAAAGATTGACTTGG	5328
QY	4608	TGCTCGCGCGCATTTTCGCCACGCGCTTCACAGGTGCTGTTCGCGCCCAAGGTGTA	4667
Db	5327	TGCTCGCGCGCATTTTCGCCACGCGCTTCACGCGAGCGGACTTGCCTCCACAGTGCACG	5368
QY	4668	CCGACGATTCGCGCGGTCTGCGCGGTGACGAAACCGCGGACACGTGTGTCCGGGTGACAGA	4727
Db	5267	CCGAGATTCGCGCGGTCTGCGCGGTGACGAGGACGCGACGTGTGTGTGTGTGTGTGTGTGT	5208
QY	4728	CCGCGTTCGCGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCGAGCTCGCGCGTGC	4787
Db	5207	CCGCGTTCGCGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCGAGCTCGCGCGTGC	5148
QY	4788	GGCGTTCGCGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCGAGCTCGCGCGTGC	4847
Db	5147	GGCGTTCGCGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCGAGCTCGCGCGTGC	5088
QY	4848	AGCACCATTTCCAGGCGCACGCTGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCG	4907
Db	5087	AGCACCATTTCCAGGCGCACGCTGCGGTTCGACGTCGAGAACCGGCTGTGATCCGACCG	5028
QY	4908	AGGTGCGCGACCCCGGCGCTCGAACTCCGCGACCGAGGCGCGCGCTGTGTGAGCCAGTTG	4967
Db	5027	AGGTGCGCGACCCCGGCGCTCGAACTCCGCGACCGAGGCGCGCGCTGTGTGAGCCAGTTG	4968
QY	4968	TTGTTCAAGGCGCCACTCCAGCGCGGCGAGAAACCGCTGCGGTTCGCGGATCTGTGCGCG	5027
Db	4967	TTGTTCAAGGCGCCACTCCAGCGCGGCGAGAAACCGCTGCGGTTCGCGGATCTGTGCGCG	4908
QY	5028	CCCAAGTGCAGGCGGTTCGAGAACCGCTTCGCGACCGCGCGAAAGATCCGACATCGGTGCGT	5087
Db	4907	CCCAAGTGCAGGCGGTTCGAGAACCGCTTCGCGACCGCGCGAAAGATCCGACATCGGTGCGT	4848
QY	5088	ACGCGCTTCAAGCGCGCTCGCGGCTTTAGACGCGCGACCGCGACGACGACGCTGCGCA	5147
Db	4847	GGCGCTTTCATCCGCGGTCTGTGTGTAGCGGCGAGGCGCACGACGATGTGTGCGCA	4788
QY	5148	GCTGATGTTTGAAGTATGAGCTGTGCGACGACGTTCCGATCTTGGCCGAAAGTCAATCC	5207
Db	4787	GCTGATGTTTGAAGTATGAGCTGTGCGACGACGTTCCGATCTTGGCCGAAAGTCAATCC	4728
QY	5208	ACCGGTGTCGCGGCGCACCGCGCTGCTCGAAGTTCGCGGCGACCTCGACGACATGTATCC	5267
Db	4727	ACCGGTGTCGCGGCGCACCGCGCTGCTCGAAGTTCGCGGCGACCTCGAATGTATGTATCC	4668
QY	5268	GGTTCCTGTTGCGGTGAAACCGCGCACCGCTCTCGAGTGCAGAGCGTCTGTAGCGGATCT	5327
Db	4667	GGTTCCTGTTGCGGTGAAACCGCGCACCGCTCTCTCGAGTGCAGATGCTGTGTGTGTGTGT	4608
QY	5328	GGGTACGCGGACGTCACAGCATGTATCGAGGTAGGTGCGGCGGTGTCTCGGCGACGT	5387
Db	4607	GGAGGCGCGGACGTTCAACACGTATGCCATATAGCGCGGTGCGGAACTCTCTCGGCGCGT	4548
QY	5388	CGGTGTATGTTTTCGCGCTGACATGACCGTCCGCGCACTCGGCGACCGTGTGTGTGTGTGT	5447
Db	4547	CGCGGTATGTTTTCGCGCTGACATGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4488
QY	5448	CGGCTTCGGTGCGAGTGCAGAGCGCGGTGAGGCGGTGCGGTGATCTCTGTGTGACGAGA	5507
Db	4487	CGGCTTCGGTGCGAGTGCAGAGCGCGGTGAGGCGGTGCGGTGATCTCTGTGTGACGAGA	4428
QY	5508	GGCGGACATTCGCGGTGTCGCGGCGAGACGCGGTTCATTCAGACCTGACCTTGC	5567
Db	4427	GGCGGACATTCGCGGTGTCGCGGCGAGACGCGGTTCATTCAGAGTGTGACCTTGC	4368
QY	5568	GGTGTCTGTGTCGACGAGACCGCGGAAATGCTGAAATCTTGTGTCTGTGTGTGTGTGTGT	5627
Db	4367	GGTGTCTGTGTCGACGAGACCGCGGAAATGCTGAAATCTTGTGTCTGTGTGTGTGTGTGT	4308
QY	5628	TGCGGTGTCGCGCGGTGACACCGCGGTGAGTGTGCGGCAAGCGCGGCGGTGTGTGTGTGT	5687

QY 2508 GCGAGTGAATGCTGGCGGCCGACGTGACGAAGTCGTGGAACGACCGAGTGCAGGCTCGGT 2567
DB 7427 GGGGCTGACGCGCTCGTTCTCCAGGACGAAGTTCTCCACACCTCGGCGAGGCTGGTGG 7368
QY 2568 CCGGGGGTGAACCGGCGCGGACGTAAGCGGGTGGCCCCCGCCAGCCCGGGGAAACCGGCT 2627
DB 7367 ACGGGGGTGAACCGGCGCGGACGTAAGCGGGTGGCCCCCGCCAGCTCCGGGAAAGCCGCT 7308
QY 2628 CCGGTTACAGGTAGACGTTCCGAGCAGAGTGAATCTGACACCGGACCTGCGGATGGGCGG 2687
DB 7307 CGCGGTAGAGGTAGACGTTCCGAGCAGAGTGAATCTGACACCGGACCTGCGGAGTGGCGG 7248
QY 2688 TCGGCGCATCTGTTCCGCGCGGATCCGCAACAGCTGGCGTGCACACCCCGACGAGGC 2747
DB 7247 TGGGACCATGCTCTCGGCGCGGATGGCGAGCAGCTCGGCGTCCACACCCCGCGCAGAGC 7188
QY 2748 TCTCCAAACCGTTAACCCAGAGTGGTCTGACATGCCCGGGGTCTCGCTCGGCGCGTGAATCGA 2807
DB 7187 TCTCCAGGGGCTAGCCCAAGGTGATGTGCAATGCGCGGGGTCTCGCTCGGCGCGTGAATCGA 7128
QY 2808 CGAACCGGGGCGAGGCGCTCCCGGCACTCGGTCCGTTGCGCCCTCGGACACACTGCGGTGCT 2867
DB 7127 CGAACCGGACCGAGCCCTCGCGCAGCTCTGTGCGCTCGAGGTCCGACAGCCGCGCTGCT 7068
QY 2868 CCGGACCGCTGTAGTCTCTCGCGGACGAGTGAAGAGTGCAGCGCGCGCGGTGCGGCTGACCT 2927
DB 7067 CGGCGCGCTGTAGTCTCTCGCGGACCGTCAAGAGTCCAGCGCGCGGTGCGGCTGACCT 7008
QY 2928 CGTTGAGTTCCGCGCATGAAATGACCAAGTTCGATGAGCGGGTGGCCCCCGCGCAGGA 2987
DB 7007 CGTTGAGTCCGCGCATGAAATGACCAAGTTCGATGAGCGGGTGGCCCCCGCGCAGGA 6948
QY 2988 TGAATGGTTGAGGCGGACCGGACCGGTCGCGCGCTTGGCGCGCATCCGCGAGGAAT 3047
DB 6947 TGAATGGTTGAGGCGGACCGGACCGGTCGCGCGCTTGGCGCGCATCCGCGAGGAAT 6888
QY 3048 CCGGAGGTTCTTCTCTGACCCGTTTGAAGCGGACCGCATGCTCGGTTGGTGGCTGTGACT 3107
DB 6887 CTTGAGGTTCTTCTTGAACGCGTTTGAAGCGGACCGCATGCTCGGTTGGTGGCTGTGACT 6828
QY 3108 CGTGGTTTCAAGCCCGTGAACAGGAGGTACGGAACCGCGCCGAGGTCCCAAGCCGCGCT 3167
DB 6827 CGTGGTTTCAAGCCCGTGAACAGGAGGTACGGAATCGCGCCGAGGTCCCAAGCCGCGCT 6768
QY 3168 GCGCGCGAGGCTCTGTTCCGTTGAGGGGAGGAGCGTTGGTGTGAAGCGTTCAGGGCGAAGC 3227
DB 6767 GCGCGTTAGCGTCTGCTCGTTGAGGGGAGGAGCGTTGGTGTGAAGCGTTCAGGGCGAAGC 6708
QY 3228 CCGGCGGGCGGCGGTGCGACACCAAGCCCGGATACCGGGTTGGTCAACCGTTTCAAGGC 3287
DB 6707 CGCGCGCGGCGGCGGTGCGACACCAAGCTCGCGGAGTGGTCAAGCGGCTTCAAGCC 6648
QY 3288 CGCGGAGAGGTACATGCGCGGTCCGCGTTCTCGACGAGCACTCGTCAATGACCGAGTGA 3347
DB 6647 CGCGCGCATGTACATCGCTTCCGCGTTGTCTGTGGGACACCTCGTCAATGACCGAGCA 6588
QY 3348 GGAATCTGTTGCGCGGACCAACCACTCGCGTGTACTTTGGACACTGTACCCGTATGC 3407
DB 6587 GCGTCTGTTGCGCGGACCAACCACTCGCGTGTACTTTGGACACTGTACCCGTATGC 6528
QY 3408 AGAAGTGAAGGAGACATGACGTCGCGGCGGGGTAGAGCCGACGCTGTACGAGGAGG 3467
DB 6527 AGAAGTGAAGGAGACATGACGTCGCGGCGGGGTAGAGCCGACGCTGTACGAGGAGG 6468
QY 3468 CGGCTCCCGCGTACCGGAGCTTCAACGCGCGCGGCGCGTTCAAGCGCGCTGATCGGT 3527
DB 6467 CAGGCTTCCGGTAGACCGCGCGGTCAAGTGGCTGCGGCGGTCAAGCGCGCTTGAATCGGT 6408
QY 3528 TGGTCAATGATTTGCGCGCGCGGCGCTTCAAGCGCGGTGCGGAGTGCAGGATGCGAT 3587
DB 6407 TGGACAGATGATTTGCGCGCGCGGCGCTTCAAGCGCGGTGCGGAGTGCAGGATGCGAGC 6348
QY 3588 CGAACAGGTGAGCAGCTCGCGAACCGGCGCGGTCGAGGTGAACCGAGCGGCGATCT 3647

DB 6347 CGAACAGGTGAGCAGCTCGCGAACCTCCGCGGCTTCAGAGCCGAGCCGCGCACGCGTCT 6288
QY 3648 GCTCCAAACGGGGTGAACCGCGAGCGCGCGTGAATGGGCGGTGAATTCGAGCGAGCCGACCG 3707
DB 6287 GCTCCAAACGGGGTGAACCGGGTGGCGCCCGTGAATGGGCGGTGAATTCGAGCGAGCCGCG 6228
QY 3708 CTTGTCTTCCGCGGTTGCGGATGCGGACCGGTGAAGCGGGTGAACCTCGGCGTGAAGC 3767
DB 6227 CTTGCAAGTGGCGGTCACCGTCAACCGGCGCGGTTGAAGCGGGAGAACCTTCGAGGCTCAGC 6168
QY 3768 CCGGACCAAGCGCGCGGCGGTGCGGCTGCGGTTGCGGTTGCGGTTGCGGTTGCGGTTGCG 3827
DB 6167 CCGGACCAAGCGCGCGGCGGTGCGGCTGCGGCGGCGGCGGTTGCGGTTGCGGTTGCGGTT 6108
QY 3828 CCGGCGCGCGGTGAAGGAGTGTCTTCAATGACGTTGCGAACCTTTCTGGCGTCTGTGGTGC 3887
DB 6107 GGGCGGTTGCGGTCGAGGTTGTCTTCAATGAGGAGCACTCTGGCACTTGTGGCGCA 6048
QY 3888 GAGGATCAAGAACCGTTGCGGTTTCCGCTGTCCACTCCGCGTTGATCAACGACCGCTG 3947
DB 6047 GGTGTCAAGAACCGTTTGGCGGTTCTTGTGATCCATTGGCGGTTGATCAAGCTGCGCTG 5988
QY 3948 GTGGTGGCGAGTGGATGACGTGCGACACCGCGCGATGTTCTTCACTGACACCGAAGGG 4007
DB 5987 GTGGCGCGGAGCGGATGATGTGCGACACCGCGCGATGTTCTTCACTGACACCGCGAGG 5928
QY 4008 CCGGTGGGGAGGAGACACCGGTTGGCGAGCACTCGGTTGTGTCAAGCGGAGCGGT 4067
DB 5927 CCGGTGGGAGCGGAGACACCGGTTGAGGAGCCCGTGAAGAGCTGTTGAGTTCTCAAGCGGAGCG 5868
QY 4068 GAGTCCGTGCGGTGAAGCGGCTGATCTGTGTGCGAGCGGAGGAGAGTGAAGGCTGTGAGC 4127
DB 5867 GAGTCCGTGCGGTGAAGCGGCTGATCTGTGTGCGAGCGGAGGAGAGTGAAGGCTGTGAGC 5808
QY 4128 ACTTCTTCGCGCGGAGATGCGCTGACAGCGGTTCAAGTTCGATGCGGTTGGCGGCTG 4187
DB 5807 ACTTCTTCGCGCGGAGATGCGCTGACAGCGGTTGAGTTCGATGCGGTTGGCGGCTG 5748
QY 4188 TCCACAGAGATGATCAAGTACTGTGATGTTGCTCTCGTGGCGGAGCGAGTGCAGC 4247
DB 5747 TCCACAGAGATGATCAAGTACTGTGATGTTGCTCTCGTGGCGGAGTGCAGC 5688
QY 4248 GTGACGCGCGGTGATGCGCGGAGCTCGTGTGTGAGAGCGGTGTTGACCCGAGTTGGC 4307
DB 5687 GATATGCGCGGAGTGCAGGAGCTCGTGTGTGAGAGCGGTGTTGAGCGGTTGGC 5628
QY 4308 TCCCTGTGTTCCGAGAACGCGTGCAGGAGGTGAAGCCCATGCGCGGCGCATCTCGCTC 4367
DB 5627 ACCCTGTGTTCCGAGAGGCGTGCAGGAGGTGAAGCCCATGCGCGGCGCATCTCGCTC 5568
QY 4368 ATCTTGGCGTTGGTGGCGGATGTCGCTGATCAACCTTGTCCGCGGAGTGGCGAGTTGGC 4427
DB 5567 ATCTTGGCGTTGGTGGCGGATGTCGCTGATCAACCTTGTGGTGGATGCCAAGTTGGC 5508
QY 4428 ATGGCCCGGATCCGTTCCGCGGAGGAGCGGCTGCGGTGACACCGCGCCGCTCGAAG 4487
DB 5507 ATGGCCCGGATCCGTTCCGCGGAGGAGCGGCTGCGGTGACACCGCGCCGCTCGAAG 5448
QY 4488 GCGGTGAACCGGCTTGTGGGCGTGAAGCTGAACCTTGGGCTGACCGGATCCGCGCAC 4547
DB 5447 GAGGTGAACCGGCTTGTGGGCGTGAAGCTGAACCTTGGGCTTGGCGAAGCCCGCAC 5388
QY 4548 GGAAGTCAACCGGCTGCTGACAGCCGAGGCGGCTGCGGTGAGAGAGAGTGAACCTTGA 4607
DB 5387 GCGCGCGCGCGGCGGTTGACAGCCGCGGCTGCGGTTGAGAGAGAGTGAACCTTGA 5328
QY 4608 TGGTGGCGGAGATCTTCCGAGCGGCTCAAGGTTGCTGCGGCTCCGAGGTTGAG 4667
DB 5327 TGGTGGCGGAGATCTTCCGAGCGGCTTCAAGGTTGCTGAGCGGAGGCTGCGGCTCCGAG 5268
QY 4668 CGGACATGCGCGGCTTGGCGGTTGAAGAGCGCGGAGCGGATGCTGAGTCAACGA 4727

Db 5267 CCGATGATCGCGCCGCTCCGCGGTGTCAACAGCAGCGGACGCTGCTCGGGGTTCAGAGCAG 5208
 Qy 4728 CCGGTCCGCGGGGTGACGTGACAGAAACCGGTGTGATCCAGACGACGTGCGCGGTGC 4787
 Db 5207 CCGGTCTCGGGGTGACGTGACAGAAACCGGTGTGACGCGCCACGACGACGTGCGCGGTGC 5148
 Qy 4788 GCGGTGACCGGAGGTGATTTGACGAGATGATCACTTCAACCGGTGACGCGCGCGC 4847
 Db 5147 GCGGTGACCGGAGGTGATTTGAGAGGATGACGACTGCGGACACGTCGTCGCGCGC 5088
 Qy 4848 AGCACCGATTGACAGGCGACGCTGTGACAGGTGAGGATGACGCTGACGCGGAC 4907
 Db 5087 AGCACCGATTGACAGGCGACGCTGTGACAGGTGACGCTGACGCGGAC 5028
 Qy 4908 AGGTGCGGACCGCGGGCTGGAATCCCGGACAGAGGCGCGCGGTGTGAGGACGATG 4967
 Db 5027 AGGTGCGGACCGCGGGCTGGAATCCCGGACAGAGGCGCGCGGTGTGAGGACGATG 4968
 Qy 4968 TTGTTCAAGGACCACTTCAGGCGGGCGAGAACCGCTGCGCGGTGCGCGATGTCGCGCG 5027
 Db 4967 TTGTTCAAGGACCACTTCAGGCGGGCGAGAACCGCTGCGCGGTGCGCGATGTCGCGCG 4908
 Qy 5028 CCGACGTGACAGGGGTGACAGAAACGCTTCGAGACCGCGGACGATGCGCAATCGCTCG 5087
 Db 4907 CCGACGTGACAGGGGTGACAGAAACGCTTCGAGACCGCGGACGATGCGCAATCGCTCG 4848
 Qy 5088 ACCGCGCTTCAAGCGGTGCGCGGTGTGAGACGCGGACGCGGACGAGCGGACGAGGTCGCA 5147
 Db 4847 GCGGTTTCATCCGCGGTGCTGCTGTGAGACGCGCGGACGCGGACGAGCATGATGCTGCGCA 4788
 Qy 5148 GCTGATGTTGACGATGCTGCTGCGGACGAGATGCTGCTGATGCTGCGGACGATGCTGCTG 5207
 Db 4787 GCTGATGTTGACGATGCTGCTGCGGACGAGATGCTGCTGATGCTGCTGCTGCTGCTG 4728
 Qy 5208 ACCGCTGTGCGGGGACCGCGGTGCTGCAAGTCCGCGGACGCTTCGACGAGCATGATACC 5267
 Db 4727 ACCGCTGTGCGGGGACCGCGGTGCTGCAAGTCCGCGGACGCTTCGATGAGCATGATACC 4668
 Qy 5268 GGTTCCTGTGCGGTGAGAACCGCGGACCGCTTCGAGATGCTGAGGCTGCTGATGCTGCTG 5327
 Db 4667 GGTTCCTGTGCGGTGAGAACCGCGGACCGCTTCGAGATGCTGAGGCTGCTGATGCTGCTG 4608
 Qy 5328 GGTTCCTGTGCGGTGAGAACCGCGGACCGCTTCGAGATGCTGAGGCTGCTGATGCTGCTG 5387
 Db 4607 GGTTCCTGTGCGGTGAGAACCGCGGACCGCTTCGAGATGCTGAGGCTGCTGATGCTGCTG 4548
 Qy 5388 CCGTGTGATGCTGCGGTGACGATGACGCTGCGCGGACGCTGCGGACGCTGCTGCTGCTG 5447
 Db 4547 CCGTGTGATGCTGCGGTGACGATGACGCTGCGCGGACGCTGCGGACGCTGCTGCTGCTG 4488
 Qy 5448 CCGCTCTGTGCGGTGACGATGACGCTGCGCGGACGCTGCGGACGCTGCTGCTGCTGCTG 5507
 Db 4487 CCGCTCTGTGCGGTGACGATGACGCTGCGCGGACGCTGCGGACGCTGCTGCTGCTGCTG 4428
 Qy 5508 GGGCGAGCATGCGGTGCTGCGGGGAGAGACGCGGTTCATCCAGGACCTGCACTGCG 5567
 Db 4427 GGGCGAGCATGCGGTGCTGCGGGGAGAGACGCGGTTCATCCAGGACCTGCACTGCG 4368
 Qy 5568 GGTGACTGTGCTGACGAGACGCGGAGATGCTGAGTACTTCTTCTGCTGCTGCTGCTG 5627
 Db 4367 GGTGACTGTGCTGACGAGACGCGGAGATGCTGAGTACTTCTTCTGCTGCTGCTGCTG 4308
 Qy 5628 TGGCGTGTGCTGCGCGGATGACCGCGGTGCTGATGCTGCGGACGCGGATGCTGCTG 5687
 Db 4307 TGGCGTGTGCTGCGCGGATGACCGCGGTGCTGATGCTGCGGACGCGGATGCTGCTG 4248
 Qy 5688 GGAAGCACTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5747
 Db 4247 GGAAGCACTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4188
 Qy 5748 GGAAGCACTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5807
 Db 4187 TGGCGGGGTGCTGCGCGGATGAGACGATGACGAGCGCGGCTGGAAGGACGCTGCG 4128

Qy 5808 GCAATCCGTGACGATACCGCGGCGGACGCGCGCGGCTTCCACCGGTGTGATGAGGCTCA 5867
 Db 4127 GCAAGCGCGGAGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4068
 Qy 5868 GCAAGCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5927
 Db 4067 GCAAGCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4008
 Qy 5928 GAGAGGTGACCCACCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5987
 Db 4007 GAGAGGTGACCCACCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3948
 Qy 5988 TGTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6047
 Db 3947 TGTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3888
 Qy 6048 GATTCGCGGTGCGGGGCGGCTGCTGCAAGTACTGATGAACTGGAACGCGGATGCTGCTGCT 6107
 Db 3887 GATTCGCGGTGCGGGGCGGCTGCTGCAAGTACTGATGAACTGGAACGCGGATGCTGCTGCT 3828
 Qy 6108 GAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6167
 Db 3827 GAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3768
 Qy 6168 TGTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6227
 Db 3767 TGTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3708
 Qy 6228 CGATCAGACCGGAGAACCGGATGCTGCGGATGGAAGATGATGATGATGATGATGATGATGATG 6287
 Db 3707 CGATCAGACCGGAGAACCGGATGCTGCGGATGGAAGATGATGATGATGATGATGATGATGATG 3648
 Qy 6288 GCCATCCGAAAGTTGCTGCGGACGCTGCAAGGCTGCAAGGAGGAGGAGGAGGAGGAGGAGG 6347
 Db 3647 GCCATCCGAAAGTTGCTGCGGACGCTGCAAGGCTGCAAGGAGGAGGAGGAGGAGGAGGAGG 3588
 Qy 6348 GCGCGAGCGGACCGCTGCTGCGGATGGAAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6407
 Db 3587 GCGCGAGCGGACCGCTGCTGCGGATGGAAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3528
 Qy 6408 GGTGCACTTCGAGCCGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6467
 Db 3527 GGTGCACTTCGAGCCGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3468
 Qy 6468 TGGCGGTGAGGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6527
 Db 3467 TGGCGGTGAGGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3408
 Qy 6528 GATGCGATGACCGCTGATGACCGGATGACCGGATGACCGGATGACCGGATGACCGGATGACCG 6587
 Db 3407 GATGCGATGACCGCTGATGACCGGATGACCGGATGACCGGATGACCGGATGACCGGATGACCG 3348
 Qy 6588 CGTTCGCGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6646
 Db 3347 CGTTCGCGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3293
 Qy 6647 GCCCGGACGCGGAGAGGCTCC 6667
 Db 3292 GGTTCAGCGGCGGAGAGGCGC 3272

RESULT 4
 AR092722/c 8051 bp DNA linear PAT 08-SEP-2000
 LOCUS AR092722
 DEFINITION Sequence 2 from patent US 5998194.
 ACCESSION AR092722
 VERSION AR092722.1 GI:10019474
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 8051)

Db	5524	GATATGCGCGCACTGCGGAGCTGTCTCGAGTGAAGACGGGTGTTAAGGGCTGTGTC	5465
Qy	4308	TCCCTGATCTCGGCGAAGCGCTGCAGGAGGTGAAGCCCATGAGCCGCGGCACTCGTCT	4367
Db	5464	ACCTCGGTCTCGGCGAAGCGCTGCAGGAGGTGAAGGCCATTCGCGCGGCGCACTCGCTTC	5405
Qy	4368	ATCTTCCGCTGTGTGTCGCACTCTCGGTACCACTTGTCTCGGCGCGATGCCAAGTTGTGC	4427
Db	5404	ATCTTCCGCTGTGTGTCGCACTCTCGGTACCACTTGTCTCGGGTGCAATCCGAAATTGTGTC	5345
Qy	4428	ATTGCGCGGATTCGTTTCGGCGCAGAGCGCGTCTGTCTGTGACACCGCCCCGCTTCGAAG	4487
Db	5344	ATTGCGCGGATTCGTTTCGGCGCAGAGCGCGTCTGTCTGTGACAGATGAGCCGCCCTTCGAAC	5285
Qy	4488	GCGGTGAACCGCTTGTGTGTCGTCGAAGGCTGAACCTTCGGGCGTACCGGATTCGCGCAC	4547
Db	5284	GAGGTGAACCGCTTGTGTGTCGTCGAAGGCTGAACCTTCGGGCGTTCGGAAGCGCGCAC	5225
Qy	4548	GGAAGTCCACCGCTGTGTGACAGCCCAAGGCGTGTGGCGGCGTGAAGAAAGCTTGAACCTTGG	4607
Db	5224	GAGCGCGCGCGCGAGTGCAGCCAGCGCGTGTGGCGGCGTGAAGAAAGCTTGAACCTTGG	5165
Qy	4608	TGTCTCGGCGGAGATCTTTCGACAGCGCTTCGACCGGAAGCGGCGCTGCGCCAGAGGTGACG	5105
Db	5164	TGTCTCGGCGGAGATCTTTCGACAGCGCTTCGACCGGAAGCGGCGCTGCGCCAGAGGTGACG	5105
Qy	4668	CCGACGATTCGCGCGGCTGTGCGGGGTGAAGAGCGCGGGAAGTGTCCGGGTTCGACAGAA	4727
Db	5104	CCGATGATTCGCGCGCGCTTCGCGGTTGTACAGCAAGCAAGTGTCTCGGGGTTCAGACAG	5045
Qy	4728	CCGGTTCGCGCGGAGTGCAGCTGCAGAAACAACCGGTGTGAGTCCGACCGAGCTTCGCGGTGC	4787
Db	5044	CCGGTCTCGGAGGTTCAGTGCAGAAACAACCGGTTTCAGGCCCAAGCTTCGCGCGGTGC	4985
Qy	4788	GCGGTGTGCGCGGAAGGTATTGAAGGCAATGATCACTTCACTCGGTGAAGTCAACCGGCCGC	4847
Db	4984	GCGGTGTGCGCGGAAGTATTGAAGGCAATGACACTTCGCGGCAACGTCCTGCGCGCGC	4925
Qy	4848	AGCAACGATTTCAAGGAGCGACGAGTGGCGTTGACAGGTGGCAATGTCGATACCCGACAC	4907
Db	4924	AGCAACGATTTGACAGGCGCAACCGTGTGAGTTGAGTGGCAGCAGTGGCGCAACCCGCGC	4865
Qy	4908	AGGTTCGCGCAACCGGCGCTTCGAACTCCCGCAACAAGGAGCGCGCGTGTGTGAGCAAGTTC	4967
Db	4864	AGGTTCGCGCAACCGGCGCTTCGAACTTCGCGCAACAAGTGTGTGTGTCGCGCAATTTG	4805
Qy	4968	TTGTTCAGAGGCCCACTCGAGCGCGGCGCAAGAACCGCTGCGGTGCGGATCTGCGCGG	5027
Db	4804	TTGTTCAGAGGCCCACTCGAGCGCGGCGCAAGAACCGCTGCGGTGCGGATCTGCGCGG	4745
Qy	5028	CCCAAGTTCAGAGGGGTGCAAGGAAGCGTCCGGGGAACCGCGCAAGTTCGCAATTCGCTGCTG	5087
Db	4744	CCCAAGTTCAGAGGGGTGCAAGGAAGCGTCCGGGGAACCGCGCAAGTTCGCAATTCGCTGCTG	4685
Qy	5088	ACGCGCTTTCACGCGGTGCACCGGTTTGAAGACGCGCGGAACGCGCAGGCGCAAGAGCTGCGCA	5147
Db	4684	GCGGCTTTCACGCGGTGCCTGTGTGTGAAGACGCGCGCAAGCGCGCAAGTGTGTGCGCA	4625
Qy	5148	GCTGGAATTTGAAGTGAAGTGTGTGTCGCGAGCAAGTTCGTTGATCTGACCGAAGTCAATTC	5207
Db	4624	GCTGGAATTTGAAGTGAAGTGTGTGTCGCGAGCAAGTGAATCTGATTCGAAGGTCATTC	4565
Qy	5208	ACCGGTGATTCGAGGAGGCAACCGCGCGTGTGCAAGTCCGAGGGAACCTTCAGACAGATGATAC	5267
Db	4644	ACCGGTGATTCGAGGAGGCAACCTGCGTGAAGTGTGGGAGGAACCTTCAGATCAGATGATAC	4505
Qy	5268	GATTCTGCTGTGCGGTGAAGACCGCCACCTTCTTCGAGTGCAGGCGTCTGTAGCGAATCT	5327
Db	4504	GATTCTGCTGTGCGGTGAAGACCGCGCGCTTCTTCGAGTGCAGGCGTCTGTAGCGAATCT	4445
Qy	5328	GAGTACCGGAGCGTCCAGCAAGTGAAGTGAAGTTCGCGGCGGTCTCTCGGAGCACT	5387

Db	4444	GCAGACCGCGCACTTTCAACAGTAGTTCACATAGACCGAGTCCGAGAACTCTCCGGGCGGT	4385
Qy	5388	CGAGTAGTATTCTCCGGCTGACAGTAGACACCGTCCGCGCAACTTCGCGAGACGTGTGGCCAC	5447
Db	4384	CGCGGTAGTTGTCCGGCTCGCGAGTAGTACCGTAGGGCCCAAGTTCGGCGACGTTCATCCCGC	4323
Qy	5448	CCGCTCTGGTGCAGAGTGCACGAGCGGTGAGAGGTGCCTGTCACTCTTTCACAGAA	5507
Db	4324	CCGCTCTGGTGCAGAGTGCACGAGCGGTGACACGTGCGCCGATGTCCTTTCACAGCA	4265
Qy	5508	GGGCGACGATCCCTGGTGTGGCGGGGAGACACCGTTCATCCAGACCTCACTCC	5567
Db	4264	GGGCGACGACCGCTGTGTGGCGGGGAGACACCGCTGTCAACAGATTTCCTCC	4205
Qy	5568	GGTGACTGTGTGTCACGAGACCGCGCAAGATGTCAAGTACTTTCCTTCCTCGATTTCGA	5627
Db	4204	GGTGTGTGTGTGTCACCGTAGCCGTACGCCGAAAGCTTGAAATCTTCTTCCTCGATTTCGA	4145
Qy	5628	TGCGGTGTGCGCGCCGAGTCCACCCGCTGCGGTTCGATGTCCGACCAACCGCGAGTGGCTT	5687
Db	4144	TGCGGTGTGCGCGCGAGTCCACCCGCTGCGGTTCGATGTCCGACGAGGGCGCGCGCT	4085
Qy	5688	GAGCGAATCTACGTTGTGTGTGCTGTCACGATTCAGCTCAGTCAAGAGGTTCATGTGT	5747
Db	4084	GACGAAATCTCCGCGCTGTGCGCGCACGTTCGAGATTCAGCTCAGCAGCGCTTTCATGTGT	4025
Qy	5748	GAGCCGAGGAGTGCAGCGCCCGACGAAAGAAAGCAGACATCGGGGTCTCCAGAGAACCCCTGGG	5807
Db	4024	TGGCCGAGGAGTGCAGCGCCCGACGAAAGAAAGCAGACATCGGGGTCTCCAGAGAACCCCTGGG	3965
Qy	5808	GCAGTCCGTTCAGTAGCCCGACCGGCGGACCCCGCTTCACACCTGTGTGATGACGACCA	5867
Db	3964	GCAAGCCGCGAGGACGACGTCGTCCGCGCCGCGGTCCGCTCAACGCGGTGAGGACGAGGCCA	3905
Qy	5868	GCACGCAACCGGAGGTCCATGTTTCACCAACGTTGTGTGACGAGACATCCCGCAGCTGGG	5927
Db	3904	GCAAGGAGTGCAGCGGTTCATGTTTCACCAAGTGTGTGTGATGATGCGCCGCGAGCTGGG	3845
Qy	5928	GAGGAGTGCAGCATCTGCAATGTGGGAGCTCCGCGCGGTCTCGAAGAACCTTCACCA	5987
Db	3844	GAGCGGTTCAGCCGGAAGTTCCGAGTGTCCGAGAGGTGTTCAGAACCTTCACGACCA	3785
Qy	5988	TGTTGCGGTTGCGTTTGCAGAGAACCAAGAACCTTCGCTCGAATCGAGGACGTTCAGCA	6047
Db	3784	TGTTCCGGTTCGTTTGCAGAGAACCAACCGCTGCTCGAATCGAGGACGTTCAGCA	3725
Qy	6048	GGAATCCGAGTGGGCGGAGTCCGTCGAAATCTCGATGAACCTGACCGCGGATTCGTTGT	6107
Db	3724	GGAATCCGAGTGGGCGGAGTCCGTCGAAATCTCGATGAACCTGACCGCGGATTCGTTGT	3665
Qy	6108	GGAACCTTCATGTAGTTGCTGCGGTGTGGCTTCGACAGGTTCGGCGAGACCTTCGACCGCTTGA	6167
Db	3664	GGAAGCCGAGTGTAGTTGCTGCGGTGTGGCTTCGACAGGTTCGGGAGACCTTCGACCGCTTGA	3605
Qy	6168	TGTTGCAAGGTTCCGACTTGGGCTGACACAGAGCGTTCAGTACGCGCTCGAACCTCCGCA	6227
Db	3604	TGTTGCCCGGCTCCGACTTGGGCTGACACAGAGCGTTCAGTACGCGCTCGAACCTCCCTTGA	3545
Qy	6228	CGATTCAGACCGAGGAACCGGATCTCGGGTTGACGATGATGGGTTGGATTCAGTCCGCTC	6287
Db	3544	CGATGAGGCGAGAGGAACCGATCTCGGGCTGACGATGATGGGCTGGATTCAGTCCGCGCC	3485
Qy	6288	GCCATTCGGAAGTTGTGTCCGACGTGACGAGGCGCTCGATGTGAGAAAGAGCGCGCGAGTCCGT	6347
Db	3484	GCCAGCGGAAGTTGTGTCCGACGTGACGAGGCGCTTCGATGTGAGAAAGAGCGCGCGAGTCCGT	3425
Qy	6348	GCGCCAGCCGACCGTCTCTCCGGGTGAAACGACGAGGTTTCATGTGTCTGAAGGGACCTC	6407
Db	3424	GCGCGAGCCGTCGCGTCTCTGGGGCTGAAACGACGAGGTTTCATGTGTGTGAACGGACCGC	3365
Qy	6408	GGTAGCACTTCGAGCCGATGCTTCGCGCGCGCGGTGGGCAACGATGTGTGACCTTCCTCGG	6467
Db	3364	GCGCCATCTTCGAGCCGCTGTCTCGCGCGCTTGTGGGGTTCAGCCGATGTGTGTTCCTCGCGG	3305

QY	3408	GAAAGTACAGCGGAAATATGACAGTCCGGGCGCGGGTATGAAACCCGACCGCTGTATGAGGAAAG	3467
Dp	6643	AGAAGTGGCAGCGGAAATGACACTCGACCCTGGGGGTATCAAGCCGACACTGTATGAGGAAAG	6584
QY	3468	CGGAGCTCCCGCCGTACCGCAGACTCTCAACCGCGCGGCGGGTTCAGCGGCGCTGATCGTGT	3527
Dp	6583	CAGGCTTGCGGTATACCGCGCGCGTTCAGTGGCGCTGGGGGTCCAGAGCGGCTTATCTGTGT	6524
QY	3528	TGCTCCAGTACTTGTACCCCGCGGGGCTTGTCTCAACCGCGGTGGCCAGCTTGGCGGATGCGAT	3587
Dp	6523	TGACCCAGTACTTGTACCCCGCGGACCGTGTCTCAACCGCGGTGCCAGAGGTCCGGGATGGGCG	6464
QY	3588	CGAACAGGTGAGGAGCGTCCCGGAAACCCCGGCGCGGTCCAGGTGGAACCGAACGACACT	3647
Dp	6463	CGAACAGGTGAGGAGGTGGCGGAATCTCCGCGCGTTCAGAGCCAGAGCCGCGACAGCTCT	6404
QY	3648	GCTTCAACCGGGGTGAAACGACAGGCGCGCTTATGTGGGCGGTGATTTCCGCGAGCGCGACCG	3707
Dp	6403	GCTTCAAGCGGGTGAACGGGTGGGCGCCCTGTATGTGGGGGAGAGGTTCGGCGAGCGCGGCGG	6344
QY	3708	CTTGTCTGTTCCGCGGTGTCCGATGCGGACCGGTTGAGGCGGGGTATCTCCGCGTTAGCG	3767
Dp	6343	CTGTACGTGGGCGTCAACCTGTCAACCGGCGCGGTGTGACGCGGAGAACCTTCGAGCTCAGCG	6284
QY	3768	CCCCACCAACGGGCGGGCGCGGTCGGGCGTGGGCGTGGGCGCGGTCCGAGATCTCCGTCCGTA	3827
Dp	6283	CCCCACCAACCGGCGCGCGCGGTGACAGTCCGGGCGCGCGCGTCCGCGAACCCCGCTTCT	6224
QY	3828	CGGCGGTTCGCGGTAGAACTGTCTTCATCGACGTGCGAACCTTCTGCGGTCTGTGTGTGC	3887
Dp	6223	GGGCGGTTCGCGTACGAGTTGTATTCATCCGGGAGCGAACCTCTTGAGACTTGTGGGCA	6164
QY	3888	GAGGATCACGAACCGTTGCGTTTCCGCTTGTCCACTCCCGGTGATTCACAGCACCGCTG	3947
Dp	6163	GGTGCTCACGAACCGTTGCGGTCTCTGTGCTCCATTGCGGTTGATTCAGCTCCCGCTG	6104
QY	3948	GTTGTGGCGGAATCGGATGACGTGACACACCGGGGGAATGTCCTCACTGGACACCGAAGGG	4007
Dp	6103	GTTGGGGGGAACCGGATATATGTGACACCGCGGAGATGTCCTGCTGGAACACCGGAGG	6044
QY	4008	CCGGTCGGGAAGGCGAGCACCCGTTCCGCGGAGCACCTCGATCTGTGTCAAGCCGAGCGGT	4067
Dp	6043	CCGGTGGGCAAGGGAGGACCCCGGTCCGAGAGCTGTTCCGATTTCTCAAGCCGAGCGGC	5984
QY	4068	GGCTCCGTGCGGTAGGGCGTCACTTGTGTGGCAGGCGCGGGAGAAATGAGGCTGTGGCAAG	4127
Dp	5983	GGCTCCGATCCGGTACGGCTGTGATCTGTGTGGCACCCCGGGGGAAGATGAGGTTGTGGACA	5924
QY	4128	ACCTTCTCCGGGCGGAGATCGCTCTGACGCGCGGTCAACGATGAGTCCGGTGGCGGTGCGG	4187
Dp	5923	ACCTTCTCCGCTCGCAGGAATGACCTTGACACTGTGTGCGGTGAGATGCCGTGTGGCGGAG	5864
QY	4188	TCCACAGAGATGATCACTACTGTGATGTTGCTCTCTGTGAGGCGGGAGCGAGTGCACG	4247
Dp	5863	TCCACAGAGATGATCACTACTGTGATGTTGTTGTGTGTGCTGCGCAGGATCGAAGCGGTGACG	5804
QY	4248	GTTAGCGCGCGTACGTGCGGCGACGCTGCTGTGTATGAGCGCGGTGTGACCCCGTTGTGC	4307
Dp	5803	GATAGCGCGGCACTTCGCGGAGCTTCGTCCGAGATGAGACCGGTGTGTGAGCGCGTTGTGC	5744
QY	4308	TCCCTGATCTCGGCGGAACGCGTCCAGGGAGGATGAGCCCATGGGCGCGGCGCACCTGCTC	4367
Dp	5743	ACCCTGTGCTTCGGCGAAGCGTTCAGGCGAGGTGAGGCCCATCGCGCGGCGCACCTGCTC	5684
QY	4368	ATCTTGCCTGTGTGTCGCGATCTCGGTGACCACTTGTCCGGGCGGATGTCGGAAGTTGTGC	4427
Dp	5683	ATCTTGCCTGTGTGTCGCGACATCCGTGACCAAGCTTGTCTCGGTGGGATCCGAAAGTTGTGC	5624
QY	4428	ATGGCCCGGATCCGTTCCGCGCACGAGCGGTGTGTGATGACACCGCGCCCGCCTCGAG	4487
Dp	5623	ATGGCGCGGATGCGGTCCGCGCACGAGCCCGTGTCTCGGTGAGCAGATGGCGCGCCCTCGAAC	5564

[illegible]

[illegible]

```

gene
CDS
3462..4634
/gene="megv"
3462..4634
/gene="megv"
/codon_start=1
/transl_table=11
/product="mycrose O-acyltransferase"
/db_xref="GI:10179844"
/translacion="MWTSTNDLDTTARBALNSLTGMRFYAAFLVFEHTVLSRLPNSVY
YADGLDFKQTTGVRGVSFFITLSGPVLTMSARASDSVMSFMRRCVKLPNHLVTAF
AAVVLPLVTQAVSGEALIPMLLIHMFPLETISFGINPVSGLAEBAFYLCFPIPL
LFMLSGIRPERLMAAIVPAIVAVVADLLPSSPPLIPGLEYSALIDWFLYTPP
ATRLSEFLIGIILARILITGWINVGLPAVLPPVFVASLPLGVYALSSSMWILP
LVLIASGATADLQOKRTFMNRVMVMGLGVSPALYVHFLVITYGADLLGFSQTEDA
PLGLALPMLIPFLAVSLVLSWLYRFLVPLPVMRWAPASARRKAPTEPQTEPRR"
4651..5775
/gene="megDII"
4651..5775
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG13910.1"
/db_xref="GI:10179845"
/translacion="MTTYVMSYLLSEYERERADILDVQKVPASGLILGQSVENFETE
YARHGIACHGVGNQTNVAKLSESVGVGDEHVTNSPAAPTVLAIDIGARPV
ADVEDVYIMDTDLVEAAVTPRTKAIIVPHLYGQCVMTALRELDRLGLVDECAQ
AHGARBRRLAIGTMSDPAAPSPFPTKUTGAGDGAAYVNTDRTABRLRYGMBE
VYVYTRPGHNSRLDEVDAELRKLRLDLYVAKRRAVQRYVDGLADLQDSGLEI
PLVPTDGNHVFYVYVVRPRDEIIKRLIDXYDLSLNTSYPMVHTMGFHLGVASG
SLPVTERRAGEVIFSLPMYPLPHLDQDRVIEAVREVITGL"
5822..6595
/gene="megDIII"
5822..6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"
/protein_id="AAG13912.1"
/db_xref="GI:10179847"
/translacion="MPNSHSTTSSTDVAERYADIYDFHGRKGTYAEADALVEVA
RKHTPQAATLLDVACGTGSHVELADSFREVGVDSLGAAMLATARNDPGRLEHGDG
RDFSLDRPDVYTCMFSTGYLVDEALDRAVANLAGLPGCTLVPEPMVPEFTRP
GMVGADIVTSGDRRISRMSTVPAGLPDRTASRMTHYTVGSPKAGIEHFEVHMTL
PARAYEQAEPQAGISCSYVGHDLFSPGLFVGVAEPRR"
6592..7197
/gene="megDIV"
6592..7197
/gene="megDIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="AAG13911.1"
/db_xref="GI:10179846"
/translacion="MRVEELIGVFTFPOTFADERGVTGAYOEDFVALGRPLF
PYAQVSTRSRGVGVGVAHTTMDPGSMAYKYCAAGPMDPAVDLRPSPTGRAPV
EYASNSVGLYTPVMKGHLFVSLIEDDTTLVYMSGYVDPKRAVHPDPELALPIPA
DLDLVMSERDRAVFTLREARDGILPDYAACRAAARVVRT"
7220..8206
/gene="megDV"
7220..8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"

```

```

gene
CDS
complement(8228..9220)
/gene="megDVII"
complement(8228..9220)
/gene="megDVII"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AAG13914.1"
/db_xref="GI:10179849"
/translacion="MGTTAGSARVVRGSRALHTSRLMLGTGVNFSGRVTDALRLMD
HALRGVNCIDTADYIGMRLYKHTREIVRGVRQAQGGREBEVLATKVGSEMERVN
DGGLSARHYAACENSIRLGVNHIDITQTHNIDRAAPMEQWQALBEHLYGSKGVY
GSSNLGWHYIAAOESBARRNLSGMTSHOCTVLAARHPELDVLPAAQAYGVGVFAMS
PLHGGLSGVLEKTLAAGTAVKSAQGAQVLLPVRVLEVEYDYCRGLADPAVGLA
WYLSRPGILGAVIGPRTPEQLDRLAAELITLGEEBIRLEIAIFAPAVDGPV"
complement(9226..10479)
/gene="megBIV"
complement(9226..10479)
/gene="megBIV"
/codon_start=1
/transl_table=11
/product="TDP-mycrose glycosyltransferase"
/protein_id="AAG13915.1"
/db_xref="GI:10179850"
/translacion="WRVLTLSFARHTHFGQVPLANALHTAGHDVRAVSPPELTDVYV
GAGTLPVLSDRLEPDISPEAAQVRYTTDDLPARGELBSWELHGEIESTSRF
VFPVYVNDLSFVDELVEFARMWRDPLVMEFTEPAGAAACAAHARLLMGSLDLY
FRRSGDRLRCORPADRPDLGMLTEVAFRFGIDYEDDAVCGMSVDQLPESFRILT
GLSVYTRTLPYNGSSVYPMQLTSDGVRVCFYGSALGITSNPFELTATLAR
FDGEIVYTRSGVRLASFPDVRVLYDPVPMNILLPGCAVYIHGGAGSMATRLHGVQ
ISVAHENDVCLRGQRTNGLAGVFLRPDEVADTLTMOALATVVEDRSHAENAEKLRQ
ALAAPTEAVVPPVLEALAHQRADR"
complement(10483..11424)
/gene="megBIV"
complement(10483..11424)
/gene="megBIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13916.1"
/db_xref="GI:10179851"
/translacion="WTHVTLIGVSGVGSALLREFTTHPLRLBAVARTSGRDPGS
AGIEHLVADLLEPVRQVAVADTVVHLVAYVAGSTWRSAAIVPFAERVNAIGMD
LVALRLARPGAPVTLFASSTQANPAAPSRVYQHKI EAERILLQATEDGVADVTIL
LPATYGSRPGSGTGRGVNTAMIRBALAGEBITWMHGGSVARNILLHYEDVATFTAL
HNHERALGVWTPSADRARPLGEIFETVVAASVARQGNPAVPVSVPPENAEANDFR
SDPDSIFERTLTGMHPRVPLABSIDRTVAALISTKE"
12181..22821
/gene="megAI"
12181..22821
/gene="megAI"
/notes="polyketide synthase"
/codon_start=1

```

Query Match 29.2%; Score 2606.8; DB 1; Length 47981;
Best Local Similarity 99.9%; Pred. No. 3.7e-177;
Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 6415 CTCGAGCCGATGCTCGGCGCGCGCGGCGCAACGTCGTGAGACGTGTCGGTGGCGCGT 6474
DB 1 CTCGAGCCGATGCTCGGCGCGCGCGGCGCAACGTCGTGAGACGTGTCGGTGGCGCGT 60
QY 6475 GGGAGGTCCGCGCGCGGCGGAGTCAAGGAACGATTGCGGATTCGCGAGTCCG 6534
DB 61 GGGAGGTCCGCGCGGCGGAGTCAAGGAACGATTGCGGATTCGCGAGTCCG 120
QY 6535 ATGACCGTTGACCCGATCCCATACGCTCTCCCGTATGTCGTGGCGGCTCGTGGC 6594

```

Db 121 ATGACCCCTTGACCCGATCCCCATACGCTCTCCCGATGATCGTGGGCGGTCCGTCCGG 180
Qy 6595 TACCGCCCGGACCTGACATTTGTGATGCAAGACCCCGCCGCACTGATGAGGCTCCGCCCGCA 6554
Db 181 TACCGCCCGGACCTGACATTTGTGATGCAAGACCCCGCCGCACTGATGAGGCTCCGCCCGCA 240
Qy 6655 CCGGAGAGAGGTCCCTGCAACAACTTCGCGGTGACCGGTCCCGCGGCTCGGTGAACCGGCG 6714
Db 241 CCGGAGAGAGGTCCCTGCAACAACTTCGCGGTGACCGGTCCCGCGGCTCGGTGAACCGGCG 300
Qy 6715 GTGAGAGACCCCGATCATTTGCTGTGTAATCTTCTACTGTGCGCGCGCACTCTTTC 6774
Db 301 GTGAGAGACCCCGATCATTTGCTGTGTAATCTTCTACTGTGCGCGCGCACTCTTTC 360
Qy 6775 TGAACCGGTGTTCCCGGTGATGACCGGTTCGCCGCCGCTGGAACCTGTGCGGTGAGACT 6834
Db 361 TGAACCGGTGTTCCCGGTGATGACCGGTTCGCCGCCGCTGGAACCTGTGCGGTGAGACT 420
Qy 6835 GACCGGTGCGCGGTGTTTCGCCGCTTCCGAACCTGCGATTCGTGATCGCGCAGGTG 6894
Db 421 GACCGGTGCGCGGTGTTTCGCCGCTTCCGAACCTGCGATTCGTGATCGCGCAGGTG 480
Qy 6895 GAGACGGGTGCTGACCGGAGATGATCTGCAATCAATGACGCTCAATGACATCTTGTAG 6954
Db 481 GAGACGGGTGCTGACCGGAGATGATCTGCAATCAATGACGCTCAATGACATCTTGTAG 540
Qy 6955 CATGTCCTCGCGCGGATCCGACAGGCGCGGAACCGCGCGCATCCGCTTTCGACGA 7014
Db 541 CATGTCCTCGCGCGGATCCGACAGGCGCGGAACCGCGCGCATCCGCTTTCGACGA 600
Qy 7015 CGTCGACATCAACGCTGACGCGGATGACACCGACGCAAGCATGCTGTGCGGACT 7074
Db 601 CGTCGACATCAACGCTGACGCGGATGACACCGACGCAAGCATGCTGTGCGGACT 660
Qy 7075 GGAAGGCTGCGCGGATCAAGGAAATGCGCTGTCACTAGACAGACGCAACAGCTGTCC 7134
Db 661 GGAAGGCTGCGCGGATCAAGGAAATGCGCTGTCACTAGACAGACGCAACAGCTGTCC 720
Qy 7135 GGGCTGCGCGGAACAGATGATCTGTGATGACGCTTCAATTTGCCCGCGCGCATCCGCTT 7194
Db 721 GGGCTGCGCGGAACAGATGATCTGTGATGACGCTTCAATTTGCCCGCGCGCATCCGCTT 780
Qy 7195 GGAATCCGTCGACCGGCTGCTGCGAGTGCAGTGCAGACCGCGGTTTGAAGACGA 7254
Db 781 GGAATCCGTCGACCGGCTGCTGCGAGTGCAGTGCAGACCGCGGTTTGAAGACGA 840
Qy 7255 GGTAGTAGCGGATGACGCGCTTCTGTCGCGCGGACGCGTGCACCTAGTGAATCCG 7314
Db 841 GGTAGTAGCGGATGACGCGCTTCTGTCGCGCGGACGCGTGCACCTAGTGAATCCG 900
Qy 7315 TCAAGCTTTCATTCGCGGAGCGTTCTATGAGCAGTTGGCGATGAGAGCGGCTGGGCGG 7374
Db 901 TCAAGCTTTCATTCGCGGAGCGTTCTATGAGCAGTTGGCGATGAGAGCGGCTGGGCGG 960
Qy 7375 GAGTTGAGATGACCGCGGCTCTCTACTGCGGAGTTGCGTGCACACCGCGCATCTGTACTCG 7434
Db 961 GAGTTGAGATGACCGCGGCTCTCTACTGCGGAGTTGCGTGCACACCGCGCATCTGTACTCG 1020
Qy 7435 ATGCTCTGTCGCGGACGCGACGACACCTTCGACCTGTGACGATGACGATGCGGCGCGC 7494
Db 1021 ATGCTCTGTCGCGGACGCGACGACACCTTCGACCTGTGACGATGACGATGCGGCGCGC 1080
Qy 7495 GGAACGCGGACCTGACGCGCAGTGCAGGCGCGGAACGTTGGGTGCTGATGACACCGACCGC 7554
Db 1081 GGAACGCGGACCTGACGCGCAGTGCAGGCGCGGAACGTTGGGTGCTGATGACACCGACCGC 1140
Qy 7555 GCGAGGTGCTGCGGACCTCGCGGCTTCAACGCGCGCGCGGACGCTGCGCGGTGAGT 7614
Db 1141 GCGAGGTGCTGCGGACCTCGCGGCTTCAACGCGCGCGGACGCTGCGCGGTGAGT 1200
Qy 7615 CAGGTGCGGACCTGCGCGGCGGCTCTGCGGCGGCGGCTTCCGCGAGTTCTAGCGCGC 7674

Db 1201 CAGGTGCGGACCTGCGCGGCGGCTCTGCGGCGGCGGCTTCCGCGAGTTCTAGCGCGC 1260
Qy 7675 ACCGAGACCGCGGCTGCGGACGATGAGACGCGGACCTGCTCCAGCAGCGGTGCGGACG 7734
Db 1261 ACCGAGACCGCGGCTGCGGACGATGAGACGCGGACCTGCTCCAGCAGCGGTGCGGACG 1320
Qy 7735 CTGATGACCGAGCTGCGGCTGCGGCTTCTGATCTGATGACGATTCGCGCGGAGGTCCG 7794
Db 1321 CTGATGACCGAGCTGCGGCTGCGGCTTCTGATCTGATGACGATTCGCGCGGAGGTCCG 1380
Qy 7795 GTGCTGCGGCTGCGGACCGCGGCTGCACTCAAGGCGGTGACCCCGACCGTCTCGGTCC 7854
Db 1381 GTGCTGCGGCTGCGGACCGCGGCTGCACTCAAGGCGGTGACCCCGACCGTCTCGGTCC 1440
Qy 7855 TGAACCTCGCGGACCGCGGCTGATGACGCGGCTGACGCGGCTGACGCGGCTGACGCGG 7914
Db 1441 TGAACCTCGCGGACCGCGGCTGATGACGCGGCTGACGCGGCTGACGCGGCTGACGCGG 1500
Qy 7915 ACCGAGACCGCGGCTGACCGGCTTCAAGATGACGCGGCTCAACCGGCGGTGCGGAGCGC 7974
Db 1501 ACCGAGACCGCGGCTGACCGGCTTCAAGATGACGCGGCTCAACCGGCGGTGCGGAGCGC 1560
Qy 7975 GCGGTGCTGCGGAGGTGCTGCGGACCTGCGGACCAACCGGTGCGGACCAACCGGTCTG 8034
Db 1561 GCGGTGCTGCGGAGGTGCTGCGGACCTGCGGACCAACCGGTGCGGACCAACCGGTCTG 1620
Qy 8035 GCGGTGCTGCGGAGGTGCTGCGGACCTGCGGACCAACCGGTGCGGACCAACCGGTCTG 8094
Db 1621 GCGGTGCTGCGGAGGTGCTGCGGACCTGCGGACCAACCGGTGCGGACCAACCGGTCTG 1680
Qy 8095 CGTGTGATGACGAGGCTGCGGACGAGTCCGCGGCTGCACTGGAACGCGGACCGGCGC 8154
Db 1681 CGTGTGATGACGAGGCTGCGGACGAGTCCGCGGCTGCACTGGAACGCGGACCGGCGC 1740
Qy 8155 GCGGTGACCGCGGAGGTGCGGAGGTGCAAGTCCGACCGGTGCGGACGATGAGTGTGTC 8214
Db 1741 GCGGTGACCGCGGAGGTGCGGAGGTGCAAGTCCGACCGGTGCGGACGATGAGTGTGTC 1800
Qy 8215 GTGCGCGGCGGACCGGTGATCCGAGGCTTCAACGATCCGACCGGTGCGGACGATGAGT 8274
Db 1801 GTGCGCGGCGGACCGGTGATCCGAGGCTTCAACGATCCGACCGGTGCGGACGATGAGT 1860
Qy 8275 CGTGTGAGGAGCGCGGATCCGAGGCTTCCGCGGCGGCGGCTGCGGCGGACGCTGAC 8334
Db 1861 CGTGTGAGGAGCGCGGATCCGAGGCTTCCGCGGCGGCGGCTGCGGCGGACGCTGAC 1920
Qy 8335 GCGGTGAGGAGCGCGGATCCGAGGCTTCCGCGGCGGCGGCTGCGGCGGACGCTGAC 8394
Db 1921 GCGGTGAGGAGCGCGGATCCGAGGCTTCCGCGGCGGCGGCTGCGGCGGACGCTGAC 1980
Qy 8395 CTGTCCGCTTCCGCGGCGGATGATGAGACGATGATCCGCTGCGGCGGCTGATGAC 8454
Db 1981 CTGTCCGCTTCCGCGGCGGATGATGAGACGATGATCCGCTGCGGCGGCTGATGAC 2040
Qy 8455 CGTTGCGGCTGAGCTGAGAGGAGAAAGATGAGGCGGCTGCTTTCATGATGCTGT 8514
Db 2041 CGTTGCGGCTGAGCTGAGAGGAGAAAGATGAGGCGGCTGCTTTCATGATGCTGT 2100
Qy 8515 CAACAGCATCTGCTGCGGCTGCTGCTGCGGACGCGCTTTCAGGCGGCGGACGACGA 8574
Db 2101 CAACAGCATCTGCTGCGGCTGCTGCTGCGGACGCGCTTTCAGGCGGCGGACGACGA 2160
Qy 8575 GGTACGCGGTGCTGCGGCTGCTGCGGCGGCTTCAACGACGATGACGCGGTGCTGACCGC 8634
Db 2161 GGTACGCGGTGCTGCGGCTGCTGCGGCGGCTTCAACGACGATGACGCGGTGCTGACCGC 2220
Qy 8635 CGTGTGCGGCTGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 8694
Db 2221 CGTGTGCGGCTGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 2280
Qy 8695 CGTGTGAGTACATGCGGACCTGAGCTGAGTGCACGAGGACCAACCAACATGCTGAGGA 8754
Db 2281 CGTGTGAGTACATGCGGACCTGAGCTGAGTGCACGAGGACCAACCAACATGCTGAGGA 2340

[illegible]

TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37812.1"
/db_xref="GI:13938933"
/translation="MNVVIGASGFLGSAVTHALADLPVRLVVARREVVSGAVDYE
THRVDLTEPGALAEVADARVPPAQIIGTSGMRISBDVVAERTVGLVRLDIAV
LSRSPHAPVVPFGSNTQVGTAVGTIDSDHDPGRVDRQKHTEQLKERTAG
AIRATSLRLPVPVPGAPGADRGVMTIRALTCQPTLMHIDGVRRELLVYDPA
ARAFTVTHADADLAGHRLGLTGRSPGLCEVTOVARSARTTGSEVSVPPPA
HMDPSLRVSEVPPARTAVTGMATVTVMAEADRVTAALAPRRAPSSPS"
complement (8228..9220)
/note="unnamed protein product; megB11-1(megDV11),
TDP-4-keto-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37813.1"
/db_xref="GI:13938934"
/translation="MGTTGASARVVGSRALHTSRMLGTVNPGRVYDDALRMD
HALRGVNCIDPADYIGMRLYKHTSELYGRWPAQGGREEVTLATKVSSESRVN
DGLSARHIVAACENSURKLGVDHIDIYQTHIDRAAPMDEWQAEHLVSGKGVY
GSSNLGSMHIAAQAESAARNLGMI SHQCLYMLAVRPELDVLPAQAVGVFAMS
PLHGILLGVLTEAGTAVKSAQRAQVLLPAVRPLVEYEDYCRRLGADPAVGLA
WVLRPGILGAVIGPRTPEOLDSDALRAELTLGEBELRELEAIFPAVDVGPV"
complement (9226..10479)
/note="unnamed protein product; megB1, mycosyl
transferase, mycrose glycosyltransferase; SEQ ID NO: 11=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37814.1"
/db_xref="GI:13938935"
/translation="MEVLLTSFARHTFQGLVPLAALHTAGDVAQSELTDDVYV
GAGLTSPFLSDRLPDISPBAQVHRVETDIDFARGBELSMWELHIBERTRF
VFPVNNDSFVDELVEFAMDMRBDLVMEPTFAGAVAAACAASRLMGLSDLGR
FRSRQDLRGORPADRPDLGMLTEVGRFGLDYSBDIAVQMSVDOLPESFRLET
GLSEVHTRTLPRNGSSVPMQLRTSDGVRCCTGCVSALGITNSQBELTATLAR
FQGEIYVTRSGLDPAVDPVNLVDVPMNITLPGCAAVIHGSGASWALHHTGVPQ
ISVAHEMDCVLRQRTAEAGVFLRPDEVDADTLQALATVEDRSHAMAKLQOE
ALAAPPAEVPVRLBALAHQHRADR"
complement (10483..11424)
/note="unnamed protein product; megB1V, TDP-hexose
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;
SEQ ID NO: 12= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37815.1"
/db_xref="GI:13938936"
/translation="MTRHYTLGVSQFVSALLRETTPLRLKRAVARTGSRDQPGS
AGIEHLRVDLLEPGRVAQVADTDVVHVAVAAGSTYMSAATVEBARVANGIRD
LVAAARPGPAPVLLPASTQANPAASRAOHKIEKRIROKATEBGVGVTLR
LPAIYHSGSPSGQGRGVVATMRALAEPLTMHSGYRRLALBEDVATFTAL
HNHEALVGVMTPSADBARPLGEIIFETVAASVAROTGNAPVAVSPPEMNEANDFR
SDPDSTBERTLGMHPRVPLAEGIDRTVAALISTGE"
12181..22821
/note="unnamed protein product; megA1, SEQ ID NO: 13=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37816.1"
/db_xref="GI:13938937"
/translation="MVDVDDLGLTRTHPPELPFMPPLCGHNEBELBARQULHAYLE
GISSEDDVAVGALARETAODGPHRAVVASVTELTALALAAQGRSPSVRVA
RPTAPVPLPGGAQMPGMAITLLAESVFAAAMRACRAPFEDVMSLTLELDPE
HLRRVVOVPAFAVQTSILAALMRSGVPAVDAGVISGIELAAABVCAVDEAARA
AALMSREVPVLRGDMMAVALSPARLAVRMDVDPAGNGRSVLTLCAPAPI
ABRVAEIAQGVRAQVYNVMSMAHSAQVAVAGKRSALTWRPGSDVPTTYAGLTG
RIJDTRELDHMRPSFRLLFVRDEATRAVLELQPGTFISSPHVLAASLQOTLVG
SPPAIVTLQRDQGGRLRFLVAQAVGTVDVWTAAPVGPGLPAPVAVETDEG
PSTEPDMAAPDHVLRARLEIVGAEETLAAGREVDAARVFRIGLDSVAVOLRTLA
TATGRDLTAMLYDHTPHALTALRGOEBERGSEETAHPTAEAPDPVAVVMAAC

Query Match 29.2%; Score 2606.8; DB 6; Length 47981;
Best Local Similarity 99.9%; Pred. No. 3.7e-177;
Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
RLPGVTSPEEFWEILLABGRDAVAGLPTDRGMDLDSLPHDPTRSGTAHORAGGFLTG
ATSPDAAPFGLSPREALAVRPOORITLSEWELEBAGIPSTLSRTNCFGVLIQO
KXGPRABRGEGVEGLMTGTTTSVAGSVAAYLGLBGRIPSTDTACSSSLVAVHLAC
OSLRREESTMALAGVTVMTPEMLVDVFSKMSLADPGRKASALADGGMAGACM
LLRLERISDARRHGPVLAVIRGTAVNSDASNLSAPNGAQRVIRQALAESGLTTH
6415 CTGAGCCGATGCTCGCGCGCGGCGGCGCAACAGTCGTGACGTCGTGCGGCGGT 6474
1 CTCGACCCGATGCTCGCGCGCGGCGGCGCAACAGTCGTGACGTCGTGCGGCGGT 60
6475 GGGAGGTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6534
61 GGGAGGTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
6535 ATGACCGTTGACCCGATGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6594
121 ATGACCGTTGACCCGATGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
6595 TACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6654
181 TACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
6655 CGGGAAGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6714
241 CGGGAAGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
6715 GTGCGAGCACCCGATGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6774
301 GTGCGAGCACCCGATGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
6775 TGAACCGGTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6834
361 TGAACCGGTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
6835 GACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6894
421 GACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
6895 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6954
481 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
6955 CATGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7014
541 CATGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
7015 CGTGCACATCACCGTGCAAGCCGCGATGACACCGACCAACCGCATGCTGTCGCGACT 7074
601 CGTGCACATCACCGTGCAAGCCGCGATGACACCGACCAACCGCATGCTGTCGCGACT 660
7075 GGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7134
661 GGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
7135 GGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7194
721 GGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
7195 GGAATCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7254
781 GGAATCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
7255 GGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7314
841 GGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
7315 TCACAGTCTTCATCCGAGCGGCTTCTATGAGAGTTCGAGTCGAGCGGCGGCGGCGG 7374
901 TCACAGTCTTCATCCGAGCGGCTTCTATGAGAGTTCGAGTCGAGCGGCGGCGGCGGCGG 960

QY 7375 GAGTTGAGATGAGCCCGGGGTCTTACTGAGGGTTCCGTGCAACGGCGATCTGTACTCG 7434
 DB 961 GAGTTGAGATGAGCCCGGGGTCTTACTGAGGGTTCCGTGCAACGGCGATCTGTACTCG 1020
 QY 7435 ATGTCCTCTGTCGGAACGGAGCAAGACCCCTGACCTGTGTACGAACGGTTGCGGGCGCC 7494
 DB 1021 ATGTCCTCTGTCGGAACGGAGCAAGACCCCTGACCTGTGTACGAACGGTTGCGGGCGCC 1080
 QY 7495 GGAAGGGGACCGGTACGCGAAGTCGGGCGGAAAGTGGGGTGGTGGTGGTACCAACGGACCGCC 7554
 DB 1081 GGAAGGGGACCGGTACGCGAAGTCGGGCGGAAAGTGGGGTGGTGGTGGTACCAACGGACCGCC 1140
 QY 7555 GCCGAGGTGCTGCGCGATCCCGGGCTTCAACCAACGGCGCGCCGACCGCTGCGCGGTGATG 7614
 DB 1141 GCCGAGGTGCTGCGCGATCCCGGGCTTCAACCAACGGCGCGCCGACCGCTGCGCGGTGATG 1200
 QY 7615 CAGGTGCGCACTGCGCGGCGGCTCTGCGGCGCGGCTTCCGAGAGTTCTACGCGCGC 7674
 DB 1201 CAGGTGCGCACTGCGCGGCGGCTCTGCGGCGCGGCTTCCGAGAGTTCTACGCGCGC 1260
 QY 7675 ACCGAGGACCGGGGCTGCGGTACGTGACCGCGACCTGGCTTCAGACGGGTGGCGCAGG 7734
 DB 1261 ACCGAGGACCGGGGCTGCGGTACGTGACCGCGACCTGGCTTCAGACGGGTGGCGCAGG 1320
 QY 7735 CTGCTGACCGAGCTGAGGTGCGCGCTTCACTGTAACGACTTCCGCGGAGAGTCCCG 7794
 DB 1321 CTGCTGACCGAGCTGAGGTGCGCGCTTCACTGTAACGACTTCCGCGGAGAGTCCCG 1380
 QY 7795 GTGCTGAGCGCTGCGGTACCGCGCGCGCACTCAAGGGCTGTGACCCCGACCGTCTCGGTC 7854
 DB 1381 GTGCTGAGCGCTGCGGTACCGCGCGCGCACTCAAGGGCTGTGACCCCGACCGTCTCGGTC 1440
 QY 7855 TGGACCTCGGAGACCGCGGTATGCTGAGCGCGCAAGTCAAGCGCGCAACAGTGGCGGTG 7914
 DB 1441 TGGACCTCGGAGACCGCGGTATGCTGAGCGCGCAAGTCAAGCGCGCAACAGTGGCGGTG 1500
 QY 7915 ACCGAAACAGGCGCTGACCGCGCTTCAACGAGATCGACCGGTGACCGCGGTGCGGACCGC 7974
 DB 1501 ACCGAAACAGGCGCTGACCGCGCTTCAACGAGATCGACCGGTGACCGCGGTGCGGACCGC 1560
 QY 7975 GCGGTGCTGTGAGGGGTGTGTGAGCGGAGCTGCGCGGCAACAGGTGGGCAACGCGGTCTTG 8034
 DB 1561 GCGGTGCTGTGAGGGGTGTGTGAGCGGAGCTGCGCGGCAACAGGTGGGCAACGCGGTCTTG 1620
 QY 8035 GCCGTCAACGAGCTTCCCGAACTGCGCGGCAAGACTTGTGCGACGACCCGAGACCGCGACCC 8094
 DB 1621 GCCGTCAACGAGCTTCCCGAACTGCGCGGCAAGACTTGTGCGACGACCCGAGACCGCGACCC 1680
 QY 8095 CGTGTGAGTGAAGAGGTGTGCGGAGCAAGTCCCGGCTGCACTTGAAGACCGCGACCGCC 8154
 DB 1681 CGTGTGAGTGAAGAGGTGTGCGGAGCAAGTCCCGGCTGCACTTGAAGACCGCGACCGCC 1740
 QY 8155 GCGTGCAGCCGCGGGGTGAGCGGGGTGACGCTCCGACCGGTGCGAGGTGACAGTGTGTC 8214
 DB 1741 GCGTGCAGCCGCGGGGTGAGCGGGGTGACGCTCCGACCGGTGCGAGGTGACAGTGTGTC 1800
 QY 8215 GTTCGCGGCGGAGAACCGGTATCCCGAGGTCTTCAACGATCCCGAACCGGTTGACAGTGAAC 8274
 DB 1801 GTTCGCGGCGGAGAACCGGTATCCCGAGGTCTTCAACGATCCCGAACCGGTTGACAGTGAAC 1860
 QY 8275 CGTGGCGGCGAGCGCGAGATCTGTGCTGCTCCGCGCGCGGCTGCGCGCGGACCGACTCGAC 8334
 DB 1861 CGTGGCGGCGAGCGCGAGATCTGTGCTGCTCCGCGCGCGGCTGCGCGCGGACCGACTCGAC 1920
 QY 8335 GCCCTGTGAGCCACCTTGGCCACCGCGGCGCTGCGGCGCGCGCGCGCGGTTGCGCGCG 8394
 DB 1921 GCCCTGTGAGCCACCTTGGCCACCGCGGCGCTGCGGCGCGCGCGCGCGGTTGCGCGCG 1980
 QY 8395 CTGTCTCCGTTCGCGGCGCGGTATCAAGAGAGTGGTCAACCGGTGCGCGCGGTGCTCAAGC 8454
 DB 1981 CTGTCTCCGTTCGCGGCGCGGTATCAAGAGAGTGGTCAACCGGTGCGCGCGGTGCTCAAGC 2040

QY 8455 CGTTGCCCGGTCAAGCTGTAGAGAGAGACGATGCGCGTGTCTTTCATCATGATGCTGT 8514
 DB 2041 CGTTGCCCGGTCAAGCTGTAGAGAGAGAGATGCGCGTGTCTTTCATCATGATGCTGT 2100
 QY 8515 CAACAGCCATCTGTTGGGGCTGGTCCGCTCGCAAGGCGCTTCCAGGGCGGCGAGACGA 8574
 DB 2101 CAACAGCCATCTGTTGGGGCTGGTCCGCTCGCAAGGCGCTTCCAGGGCGGCGAGACGA 2160
 QY 8575 GGTACGGGTGTCGCTCGCGCGGCTTCAACCGAGCGTCAACCGGTGCGGTCTGACCGC 8634
 DB 2161 GGTACGGGTGTCGCTCGCGCGGCTTCAACCGAGCGTCAACCGGTGCGGTCTGACCGC 2220
 QY 8635 CGTCCCGGTGTCGAGCACTGTGAACCTTGTGAGTGGCAAGCCACCGCGGCGAGACAT 8694
 DB 2221 CGTCCCGGTGTCGAGCACTGTGTGAGTGGCAAGCCACCGCGGCGAGACAT 2280
 QY 8695 CGTGAATTACATGGGACCCCTCGACTGGGTGACCAAGGCGACCAACATGTCTGAGGA 8754
 DB 2281 CGTGAATTACATGGGACCCCTCGACTGGGTGACCAAGGCGACCAACATGTCTGAGGA 2340
 QY 8755 CGACCTCTGAGGCAATGAGACCAACCTTCAACCCGACCTTCTTCCGCTGTATGACCCCGA 8814
 DB 2341 CGACCTCTGAGGCAATGAGACCAACCTTCAACCCGACCTTCTTCCGCTGTATGACCCCGA 2400
 QY 8815 CTGCTCATGACGCGGATGCTGAGTTCGCGCTCTTGGCGTCCGACCTGATGCTGTG 8874
 DB 2401 CTGCTCATGACGCGGATGCTGAGTTCGCGCTCTTGGCGTCCGACCTGATGCTGTG 2460
 QY 8875 GGAAGCGCTGACCTTCCGCGCGCGCGGATCGGGGCGCGGATCAACGGAAACCGCGACCGCCG 8934
 DB 2461 GGAAGCGCTGACCTTCCGCGCGCGCGGATCGGGGCGCGGATCAACGGAAACCGCGACCGCCG 2500
 QY 8935 GATCTGTGAGGTTCGAGACGTGCGCACCCGCGCCCGGACAGACTTCTGTGCACTGTGCGC 8994
 DB 2521 GATCTGTGAGGTTCGAGACGTGCGCACCCGCGCCCGGACAGACTTCTGTGCACTGTGCGC 2580
 QY 8995 CCACCAAGAGGTGAGCAACCGGAGAGATCC 9024
 DB 2581 CCACCAAGAGGTGAGCAACCGGAGAGATCC 2610

RESULT 8
 AY623658
 LOCUS
 DEFINITION
 Aeromicrobium erythreum putative transcripional repressor and putative dehydrogenase/reductase genes, complete cds; erythromycin biosynthesis gene cluster, complete sequence; putative oxidoreductase and lipW genes, complete cds; monomine oxidase gene, partial cds; and unknown gene.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Aeromicrobium erythreum
 Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Propionibacteriaceae; Nocardioidaceae; Aeromicrobium.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 Submitted (14-FEB-2005) Fernald, Inc., W. Campbell Park Drive, Chicago, IL 60612, USA
 Sequence update by submitter

0Y	6226	CACATCGAGACCGAGAGAACCCGATCTCGGGTTGGACGATGATGGATTGCATCGATCGCG	6285
0Y	9315	CACAGACGAGGGCGAGAGAACCCGACTCTCGGCTGACGAGATCGGCTCCACCACTCGCG	9374
0Y	6286	TCGCCATTCGAAGTTGGTCGCGACGCTGAGGAGCCCTCGATGAGAGAGAACGCCCGAGTC	6345
0Y	9375	CCGCGACACCGGTGGTTGCTCGTAGACGGCCCAACCCCTCGACCGAGAGAACCGGCCGAAATC	9434
0Y	6346	GTGGCGGACCGGACCGGATCCCTCCGGGTGGAACGACAGAGTTTCATGTGCTGAAGGGCAC	6405
0Y	9435	GTGCGCGATTCCTCCGTGCGTGGCCGCTCCAGTCTCAAGGTCTGGGAAACGAGGT	9494
0Y	6406	TCGGTGCACCTCGAGCCGATGCTCGGCGCGCGCGGAGGACCAACGATGCTGACGTCGTC	6465
0Y	9495	CCGCTCGACCTCGAGGTGCGGACGCGCGGACCGCTCGGCGAGCCAGCCGAGGAGGCTC	9554
0Y	6466	G 6466	
0Y	9555	G 9555	
RESULT 9			
AB088224/c		210614 bp DNA linear BCT 11-VUN-2003	
LOCUS	AB088224	210614 bp DNA linear BCT 11-VUN-2003	
DEFINITION	Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.		
ACCESSION	AB088224		
VERSION	AB088224.1 GI:30698345		
KEYWORDS			
SOURCE	Streptomyces rochei		
ORGANISM	Streptomyces rochei		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinometales;		
AUTHORS	Streptomyces rochei		
TITLE	Streptomyces rochei		
0Y	1	Kinaeshi, H., Fujii, S., Hatani, A., Kurokawa, T. and Shinkawa, H.	
0Y	2	Physical mapping of the linear plasmid pSLA2-L and localization of	
0Y	3	the eryAI and actI homologs	
0Y	4	Biosci. Biotech. Biochem. 62, 1892-1897 (1998)	
JOURNAL			
REFERENCE			
AUTHORS	2	Suwa, M., Sugino, H., Sasooka, A., Mori, E., Fujii, S., Shinkawa, H.,	
TITLE	3	Nimi, O. and Kinaeshi, H.	
0Y	4	Identification of two polypeptide synthase gene clusters on the	
0Y	5	linear plasmid pSLA2-L in Streptomyces rochei	
0Y	6	Gene 246 (1-2), 123-131 (2000)	
JOURNAL			
PUBMED	10767533		
REFERENCE			
AUTHORS	3	Hiratsu, K., Mochizuki, S. and Kinaeshi, H.	
TITLE	4	Cloning and analysis of the replication origin and the telomeres of	
0Y	5	the large linear plasmid pSLA2-L in Streptomyces rochei	
0Y	6	Mol. Gen. Genet. 263 (6), 1015-1021 (2000)	
JOURNAL			
PUBMED	10954087		
REFERENCE			
AUTHORS	4	Mochizuki, S., Hiratsu, K., Suwa, M., Ithi, T., Sugino, F., Yamada, K.	
TITLE	5	and Kinaeshi, H.	
0Y	6	The large linear plasmid pSLA2-L of Streptomyces rochei has an	
0Y	7	unusually condensed gene organization for secondary metabolism	
0Y	8	Mol. Microbiol. 48 (6), 1501-1510 (2003)	
JOURNAL			
PUBMED	12791134		
REFERENCE			
AUTHORS	5	(bases 1 to 210614)	
TITLE	6	Mochizuki, S., Hiratsu, K. and Kinaeshi, H.	
0Y	7	Direct Submission	
0Y	8	Submitted (15-JUN-2002) Haruyasu Kinaeshi, Hiroshima University,	
0Y	9	Department of Molecular Biotechnology, Graduate School of Advanced	
0Y	10	Sciences of Matter, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima	
0Y	11	739-8530, Japan (E-mail: kinaeshi@hiroshima-u.ac.jp,	
0Y	12	Tel:81-824-24-7869, Fax:81-824-24-7869)	
0Y	13	The nucleotide sequence has been determined by using restriction	
0Y	14	fragments and nested deletion fragments of the ordered cosmid	
0Y	15	library of pSLA2-L.	
0Y	16	pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries	
0Y	17	143 ORFs.	
0Y	18	Gene prediction was based on the unique codon usage in Streptomyces	
0Y	19	(Bibb et al., Gene 30:157-66 (1984)) using the FramePlot program of	

Isihikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as implemented at <http://www.nih.go.jp/~junc/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtc, cgg or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polypeptide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

Location/Qualifiers

1. 210614

/organism="Streptomyces rochei"

/mol_type="Genomic DNA"

/strain="7434AN4"

/db_xref="taxon:1928"

/plasmid="pSLA2-L"

/note="linear plasmid"

1. 1992

/note="left terminal inverted repeat, TIR-L; shows 99.4 % (1981/1992) sequence identity to TIR-R (complement (208623..210614))"

683..2188

/note="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L

ORF1 (501 aa)

similar to AL590463 Streptomyces coelicolor putative helicase, SCPI_135 (879 aa); homology is seen until the inner end of TIR-L"

/codon_start=1

/transl_table=11

/product="putative helicase"

/protein_id="BAC76459.1"

/db_xref="GI:30698346"

/translation="MSTTSRTDQREBAQGEAVDVRALELPRLAPERGLRQVIMATGTTTVEARSAKELAGVNLVPSLDLITTEAMRARGTGMIGVSLRGESVAFPTTQTVLWVRPFDFKVTAFATASGLGLELRAHRCGLPGMDLIYDPAHRIKPGHRAVRLIGFTSGIATDGTVEKGLGSVAVGEGVTRCDSVYADVRSMPDVOAVARARILIPGGKASVILVPIVLDPEGETADMILTSRPYNGLALFRQAPLITGHGEGFAAFPGFEVSPVIGVGAQGSAGASRSITPSAMSLISWRAP"

complement(2315..3595)

/note="ORF2 (426 aa)

similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)"

/codon_start=1

/transl_table=11

/product="putative secreted protein"

/protein_id="BAC76460.1"

/db_xref="GI:30698347"

/translation="MRRKRALILGIPIGLVAVLAPAPAHASADTVYGTAAPEPTAPADSTVEDQTLRQTVHLSVAGSLIRFTNEGTSLITGEVNAHAPAOGATATADGTRVAVRFGGRPATLAPGTRMSHSDVALPTTAGGLVSLVLPRTPSIVSAASALGKFDVATVGGPARAVTVLLGNDVGGIAPASDVTAEELTAGKQLIBRAHBLKFDICATVITPAGDTIGVFTTPRRRVRQVNDWMTSGAFTVLDFAVLDPADPMLIPATDGGGLPNDAGMAAMARAFPLSLR"

complement(3659..4645)

/note="ORF3 (328 aa)

similar to Y00459-2 Streptomyces griseus regulatory protein, Scrr (350 aa)"

/codon_start=1

/transl_table=11

/product="Scrr-like regulatory protein"

/protein_id="BAC76461.1"

/db_xref="GI:30698348"

/translation="MTAALAGNIIDQETRSIPISLVGSPRLCGEDAEHVAIADTD
EALPRTIVERTSRTVIDGKRVLAAXKAGHTTEVLRFQGAABEAFLIARSMNTGL
PLSROBBAARIIAOWPHLSRAVAGIAGIKATYARALPAAGTAPPOARBGD
GRIAPIDGTGRKAABELLQRPQASREYARAGISIPATASVRRRLASGRVPER
NTGARPAAGTGRATPGGAPEVSVVPIRPEDSPVLRLRDSPLHKSGRLL
LRLIQCAGVERTLMAAQTPPHCTDVLAELEBYADLAEAPREVARVDG"
complement (4922 .5812)
/note="ORF4 (296 aa), lankacidin biosynthesis protein
similar to AE004625-4 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein B (304 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76462.1"
/db_xref="GI:30698349"
/translation="MTLITRAGAGFPFNNCAALCARGRBELPARSGCAVSGDG
RDMWLNASPDITQLLAAPALTPGPRTIPVRGVLLTAEVDHLAGLVAGTGL
TVYAAPVRGALSAELPVRGLDRVAPWMDRATAPAGVAVAGLTATAPVTAKK
YAHAPDPAWVCAVRIEDPATGALVYACLATMPDGPFDLASATCALDGFPSA
GELGTATSSAGOSLMGHLFVAPGSGSLAALRHGRLRIYTHIANTNPLDPSGA
HAVERAGVEVLPDSELYT"
complement (5815 .6909)
/note="ORF5 (364 aa), lankacidin biosynthesis protein
similar to AJ277117-5 Gluconobacter oxydans putative pqsB
protein (359 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein E"
/protein_id="BAC76463.1"
/db_xref="GI:30698350"
/translation="MADPAPAGAPMLIELTHRCPLHCPYCSNPLIEVREAEITCEQ
MTDILTOARELGVOMHPSGSEPLARPDLDVIGHARRGAYVNIATSGVGLAEBAH
DIARCGDVHOLSLDAPDPAAGQALAGARVHTAKLAAARVTAAGPIUTNYLHGN
IDRTGKWDIADLADGRILEANTQYTGMLRNALAPPAALAEAREVNRARY
AGSELVYAADYDIDRKPCCMDGWSITDITVPADVDLPCPAVIAITLIPVNAAR
PLSEIVASRSEFNAVGTGWMREPCRTCEHRADHGCGCOAQLTGDAATADPACGL
SPHRSIVDALAEITDGPVAPFVRGSPVPA"
complement (6902 .7174)
/note="ORF6 (90 aa), lankacidin biosynthesis protein
similar to AL603642-200 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein D (98 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translation="MTGLPEPTVPRLPQVRLTRDPARGELALLPERVVLNDTAAY
LAHCDGTSLAGIVERLAESEYGAEDVRELLRLAQRRVVDLHG"
complement (7171 .7899)
/note="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translation="MSMSVTRVDAIPWSEAEFRQLHALLESSYWDHPPHRRMEGL
DSGELMLANRWYQRCBQKDAIVANCPLEVRQWMSRVYHDGADACGGAEX
WTLAEAVGIRDEVDERVLATGATVADAVDPAARRPMLAEAGTLEPSGL
AHLGLRHEHPWIAEGEFYFARIEVVGPEGRSLDLVARRAVSREQOACVRAIA
PCRVINAVLDSIDYHTNGATRS"
complement (7926 .8030)
/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="MRTSGKEILPAKKAMHRPDEVITIDGMEVTAYPEFR"
8420 .10057

/note="ORF9 (545 aa), possible lankacidin resistance
protein
similar to AL133424-19 Streptomyces coelicolor putative
ABC transporter ATP-binding protein (544 aa)"
/codon_start=1
Query Match 16.9%; Score 1513.6; DB 1; Length 210614;
Best Local Similarity 59.5%; Pred. No. 6.9e-100;
Matches 2569; Conservative 0; Mismatches 1736; Indels 13; Gaps 5;
1453 GCCAGCTGTGGGGGTGCCCCGAGATGACCAAGACAGTTTCGTGTCGCGGCC 1512
92581 GACCTGTGTGTGGAGCGGTACAGTTTCGCGGGGCCATTCGCGGGGGGTGGGCG 92522
1513 CTGTGACATTCAGATGACAGACCCGACCGATCCGACCTGTGCAAGCATCAG 1512
92521 GCGGCCACGCGCTGTGTGTCTGCGCGGAGACCTTAAC--GCTTACTTCCCGGACG 92465
1573 GTGCTGAACCGGTCACTTCTTCACTGTCTGACAGAGTGTGGGAAGCGGGCGACCC 1632
92464 TTCTTGACCGGCTTGGCTGAGACAGCCCGCTGACGCGGCCCACTGCTGCGAGTTGG 92405
1633 CGGACGACCTGATCTCCGGCTGTGTGTGCTGCGCGAGGTGACCGGGCGACCTCGAGAC 1692
92404 CTGGCGGACGCGCGCGGGGGCCACGCGGGGTGCGTTGACAGAGAGCTGTGTGCGCCAG 92345
1693 GTGAGGCGGCAACTTCTTCCACAGGCTTGTCTGCGGGGACATCAACCAACCGTC 1752
92344 TGGAGCGTGACCTGTGTGCTGCGCGCGCGGCTTCGCGCTGACCTGTGCTGCGAGTGGCCG 92285
1753 CTGCTGGGCAACATCGTCGCGACCCCTGACAGACACCGGAGATCGAGACGCGCGCGCC 1812
92284 CTGCGGATGTGTCGCTTCAACAGCGCGCGCGCTGTGTGCGGACCTGTGCGCAAGCGCGGG 92225
1813 GAGAACCGGAGTCTGATGCGCGCGGATCATCGAGAGATGTGTGCTTCCGCGCGCGCTTC 1872
92224 CCGCGCGCGCGGTGCGTGAAGAGCGGGGTCCACGCGGGACCGGCTTCAACCGCGACCG 92165
1873 CCCAGATGAGCGGACCAAGACCAAGGCGCACACCGTGTGGGGTCAAGATCCGCGCC 1932
92164 GCGCTTTCACCGTTACTCTCGGGGTGTGCGGAGGTGACGCGGAGATCGTCATCC 92105
1933 GACGATGATGATCAACCTGTGGGTCTGTGGGCCAACCGGATCTCCCTGGCGATCCCGAC 1992
92104 GCGGACCGCGCGCGCTGTGGGGGTGTGCGCGCTTCGAGCAACCTCCGGTGTGTGCGC 92045
1993 CCGGACAGTGTGACCCGCTCCGCAAGATCGGTGTGCGCGCGAGCTCTTCGGGAC 2052
92044 TTGCGCCCGCTGAGCGTCTGTGTCAGAGCGTGTGCGCGCGGTGTCACACGCTGCGCTTC 91985
2053 GCGGTGCACTTCTGTGTGAGTCCCGCTGTGCGCGCGCTGTGAGAACAGAGTCCCTGTGAG 2112
91984 GGCACCTGTGACGCGCGCTGTGACCAAGGAGTGTGCGGAGTCAATGTGCTCATCTGG 91925
2113 GAGATATGCGCCCGGTGACGTGTGAGTGTGCGCTGTGACCGGAGACAGACAGCTGTGCTAC 2172
91924 GACACCGTCTGTGGGGCGGACAGACGCGCGCGCGCGCGCTGTTCCTGCGGAG 91865
2173 TTGCAACAGATGCTCTCGGACCGCGGACCTTCCCGGTCTGTGCGCGCGCTGACCCCGGCG 2232
91864 GCGGAGGTACCGCGGAGGCGCTGTGGGAGCAAGCGTGTGCGCGGTGCTGCCCAACCTCTC 91805
2233 GAGTCCGCTGAAC--CCTTGCGCTCGACCGCGCGGNNNNNNNNNNNNNNNNNNNNNN 2290
91804 TTCTTCGACGGGGCGCGCGCGGTGCGGGCGAGAGCGCGCGGAGACGCCCAAGAC 91745
2291 NNN 2350
91744 GTGTGTCCGCTGTGAGGAATCAACCGGACGACAGGCGCGCGCTGCTCCGTGTGCGGCG 91685
2351 CACNCCNCCGCGCGCGCGGGGGGTGTGACGCGCTTCAAGACGCGCGGATCAAGGCCCGCA 2410
91684 GCGGCGGGGTGAGCGCGGGGACGCTCAAGCTGAGCGGTCCGCGCCCAAGAGGCCCGCG 91625

QY 2411 TGCTGATCCACCCCGTGGGAGAGCTCCGTTGAGTTGTTAGCCGGGGGTGACCGAG 2470
DB 91624 TACCCCCCAAGCGGTGGAGAGCTCCGTTGAGTTGTTAGCCGGGGGTGACCGAG 91565
QY 2471 TGGTGGAAACCGTGGAGAAAGACTCGTCCCGCGTGGCGAGTGAATGCTGGCGAG 2530
DB 91564 TGGTGGAAATCGTGGAGAAAGACTCGTCCCGCGCGCGCTGAGTATGCGAGTTGCTC 91505
QY 2531 GTGACGAAGTGTGGAGAGACCGAAGTGCAGGCTTCGGGTCCGGGGTGAACCCGCGAG 2590
DB 91504 TCGACGAACGCGCGAGACGACTCGTGCAGGCTTCGTCCGGGGTGAACCGGTCCGAGATG 91445
QY 2591 TAGGGGGTGGCCCCCGGAGCGCGGGGAAAACCGGCTCCGGTTACAGGTAGACGTTCCG 2650
DB 91444 TAGGGGTCCGGCGCGCGAGCGAGCGAGGAGCCCACTCGCGGTGAGGTACACGTTCCG 91385
QY 2651 AGCAGTTCGATCTGACCCGACACTGCGAGATGGGCGGTCCGAGCGCATCGTTCCGCGCG 2710
DB 91384 AAAAGTTCGACTGACCGGAGCTGCGGGGTGGCGGAAACCGCGCATGGCGCGGGGTG 91325
QY 2711 ATCCGCAACAGCTGGGCGTGCACACCCGAGCGAGGCTTCGAAAGCTTACCCAGGTG 2770
DB 91324 ATGCGCGGACGCGGCGTTCACCCCGCTGGCAGGCTTCAGGGCGTACCCAGAGTGC 91265
QY 2771 GTCGCAATGCCCGGGGTCCGCTCGGCGGGGTGATGTCAGAAACCGGGAGAGGCGCTCCG 2830
DB 91264 ATGTGACGAGAGGGGTCCGCTCCCGCGGTACGAGTGAAGTCCCGAGCGCTTCGCGC 91205
QY 2831 AACTCGATCCGTTGCGCTCCGACAACTGCGGTGTCGACCGCTGATGCTCCGCG 2890
DB 91204 AGCGGGGGCGCTCGTCCGGCGGAGCGCGCGGTGTCGCGCGCTGATGCTCCGCG 91145
QY 2891 ACGGTGACGAAGTGCAGCGCGCGGTGCGGGCTGGACTCGTTGAGTTCCGCGATGAAGTGC 2950
DB 91144 ACCGTCAAGAACTGAGAGAGAGCGGCGGGGAGCGCTGTCAGGCGCGGACGAAAGTCC 91085
QY 2951 ACCAGTTCGATGAGGCGGTTGGCCCGCGCGGAGGATGATGTGGTTGAGGCGGAGCGG 3010
DB 91084 ACGAGTTCATGAGCGGTGCGCGCGCGCGGAGGAGATGATGATGATGAGGCGGAGCGG 91025
QY 3011 ACCGTGCGCGCGCTTCGCGCGGACATCCGACGAAGTCCCGAGAGTTCTTCGACCGCT 3070
DB 91024 GTCCGGGGCGCGTGTCCGCGCGCTTCGACATGTAAGCGGCCAGGTTCTCCGTAACCGC 90965
QY 3071 TCGAAGCGCCACCGCTTGCCTGGGTGCTGATGCTGCTGTTCAAGCCGTAACAG 3130
DB 90964 CCGAAGCGCGCGGCTTGTGCTGTGGTGTCTGATCTCTCTGCTCAGCCCGTAACAGC 90905
QY 3131 GAGGTACGGAACCGGCGCGAGGTCCGACAGCGCGGCTGGCGGCGGAGGCTGATCGGTG 3190
DB 90904 GAGGTGCGGATGCGGTGACGACACAGCCCGGCTGGCGCTCCAGGCTGGGTG 90845
QY 3191 AGGGGGAAGCGTGTGTGATGACGGTCAAGGCGGAAGCCCGCGCGGCGGTGACAC 3250
DB 90844 AGCGGTGAGCGTTGTGTGACGCTGATGTAAGGCGCGCGCGCGCGCGCGCG 90785
QY 3251 AACGCGCGGATACCGGGTGTGCAACGTTCAAGCGCGCGGAGAGGTAATGCGCGCTC 3310
DB 90784 AGTTGCGCGAGCCCGGATGTGTGAGCGGTTCACCGCGCGGAGAAAGTACGCTGAG 90725
QY 3311 GGGTCTCCGAGCGGACCTGCGTGAATGACCGGAGTGAAGGATCTCGTTGCGCGGACAC 3370
DB 90724 GCGTGTCTCCGAGCACTGCGTGAATGAGAGCGGAACATGACGCTTGAAC 90665
QY 3371 GACTCGCGGTGCTGATCTTGGACCTGTACCCGTAAGAGTGAACAGCGGAACATGAC 3430
DB 90664 GCGGAGTGTCTGTAAGCGCGCGCGGTGACGCGCACAGAAATGGAAGGAAACATGAC 90605
QY 3431 GTCCGCGCGGGTGAAGCCCGACGCTGTACGGAAGCGGGCTCCCGCGTACCGGAGCC 3490
DB 90604 GTCCGCGCGGATACGACCGAGCTGTGAGGGAAGGCGCGGCGCGGTGACCGCGCG 90545

QY 3491 TCCACGCGCGCGCGGCTTCAAGCGCGCTGATGCTGTGTTCTGACGATCTTGCCCGCGG 3550
DB 90544 TCGAGGCGCGCGCGCTCAAGCGGAGGATGTGTTGATCACTTCCCTGCGGG 90485
QY 3551 CCTGTCTCAACCGCGGTGCGAGTGCAGTGCAGTGAACAGGTGAGACAGTCCGCG 3610
DB 90484 CCGCGTGCACGCGCGGTGCTGACTCGGTGACCGGTGAGAAAGAACCGAGCGAGCG 90425
QY 3611 AACCGGCGCGGTGCTCAAGTGAACCGACGAGCTGCTCAACCGGGGTGAACGAGAG 3670
DB 90424 AAGCGCGGCGGTGCTCAAGCGGAGGATGAGCGCGCTTTCAGAGGTGTGAACGAGT 90365
QY 3671 GCGCGGTGATGGCGGTGAGTTCCGCGAGCGCGAGCGCGTGTGTTGCGCGGTGCGAT 3730
DB 90364 GAGCGGTACCGCGGTGCGAGGATGACGAGGTGCTCCGCGGTGCGCGCGCGCGCG 90307
QY 3731 GCGGACCGGTGAGGCGGGTGAACCTGCGCGCTGAGCGCGCGGACCGAGCGCGGTG 3790
DB 90306 -CGGTGTCCGCGAGCTGGCGGAGGATGCGACCGCGCTGATGCGCGCTGACGAGT 90248
QY 3791 GCGTGGGTGATGCGCGGTGCGAGTCTCCGCTGATGCGCGGTGCGGTGAGTGTG 3850
DB 90247 CTTCCGAGGCGGTGCGCGGTCTGTATGCGGGGCGGGCGGTGCGGTGAGTGTG 90191
QY 3851 TTGATGACGTTGCGAACCTTTCGCGGTGCTGTGTGCGAGGATCAAGACCGTTGCTT 3910
DB 90190 -TCCAGGAACTGTATCTCCGCGTGCAGCTGGGAGACGAGCGCGCTCAGAGAGCG 90132
QY 3911 CCGCTTGTCCGACTCCGCGTGAATCAACGACCGCTGTGTGTGCGAGTGGAGTGAAGT 3970
DB 90131 TGTGTGGCGCCATGCGATGATGTCGCGCGCGCGCTCCGCGGAGCGGAGTGTCT 90072
QY 3971 GCACACCGCGGAGTGTCTTCACTGAGACACCGAGGCGCGGTGCGGAGGAGGACACCG 4030
DB 90071 GCACACCGCGGAGTGTCTTGAAGAGACCGAGCGCGGTGCGGAGGAGGATGATCT 90012
QY 4031 TTCGCGAGCACTCGGTGTGTGCAACCGAGCGGTGCTCGGTGCGGTAGGCGGTAT 4090
DB 90011 CTTCCGAGAGCGGTGCGGTGTTCCAGCGCGAGCGGGGAGTGGTCCGTGAGCGGAT 89952
QY 4091 CTGGTGGAGCGGGGAGGAGTGAAGGTGTGTGCAACGACCTTCTCCGCGCGAGATGCG 4150
DB 89951 CAGATGCGAGGGGTGAGGATGACGCTGGGGGAGAACCTTTCGCGCTGACAGAGG 89892
QY 4151 CTGACGCGGTGACGGTGATGCGGTGCGGAGCGGTGCGGTGCGGTGCGGTGCGGT 4210
DB 89891 CCGAGCGGTGCGGAGTGAACCGGTGACCGGTGACCGGTGCGGTGCGGTGCGGT 89832
QY 4211 GTAGTTGCTCTCTGTCGCGGAGGAGGATGACGGTGACGCGGTGCGGTGCGGT 4270
DB 89831 GTGGTTGTGCGTGTGCTGCTGTGGAAGAGACCTCGGTGCGGGAGCGCGGTGAG 89772
QY 4271 CTGCGGTGTGAGAGCGGTGTTGACCCGTTGTGCTCTGTTGTTGCGGAGAGCGGT 4330
DB 89771 TTCCGCGGTGAGAGGGGTGTTCTCAAGATGCGCTGCGGGTGCAGGAGGCTC 89712
QY 4331 GAGGAGGTGAGCCCATGAGCGCGGAGCACTGCTCATTTGCGGTGAGTGCATCTC 4390
DB 89711 GAGGAGGTGACACCATGAGCGCGGAGCACTGCTCATTTGCGGTGAGTGCATCTC 89652
QY 4391 GGTGACCACTTGTCCGCGCGATGCGAGTGTGCAATGAGCCCGGTGCGGAG 4450
DB 89651 GGTGACGAGCTTGTCCGGTGAGTGCAGAGTGTGCGAGGCGACAGCCGTTGCGGAG 89592
QY 4451 CAGGCGGTGCTGAGTGCACCGCGCGCGCTTCAAGGCGGTGACCGGCTTGTGCGGT 4510
DB 89591 TTCCGCGGTGCTGAGTGCACCGCGCGCGCTTCAAGGCGGTGACCGGCTTGTGCGGT 89532
QY 4511 GAGGTGAAACCTTCGCGGTGACCGGAGTGCAGCGGAGCGGTGACCGGCTTGTGCGGT 4570
DB 89531 GAGGTGAAACCTTCGCGGTGACCGGAGTGCAGCGGAGCGGTGACCGGCTTGTGCGGT 89472
QY 4571 CAGGCGGTGCGGTGAGGAGGCTTGAACCTGTGTGCGGCGGAGTCTTTCGCGAG 4630

Db	89471	GAAAGCGTGGGCGCGGTGCGAAGAACAGTCCGACGCCGTGCTCCGCGGCCACGTTCTCCAG	89412
Qy	4631	CGCTTCACAGTGTGTGTGCGGCCCCACAGGTGTACCGACGATCCGCGCGGTCTGCGG	4690
Db	89411	CTCCAGAGACCGGGGCGCGCGCCGCCCCACAGGTGACGCGCAGAGATGGCCCCGCGTCTCGGG	89355
Qy	4631	GGTGACGAGCGCGCGGACAGTGTGTCCGGGTGACACAGACCGGTGCGCGGTCTGACGTTCGA	4750
Db	89351	GGTGTATCAGCTTCTTCGACCCCGGCGGGGTGAGGCAACCCGTGTACCGGGGTGACGTTCGCA	89292
Qy	4751	GAAACACCGGTGTGAGTCCGAGCCAGCTCCGCGCGGTGCGCGGTGCGCGAGAGGTGATTTGA	4810
Db	89291	GAAACACCGGTGTGAGGCCCCACCGCTCCGCGGTGCGCGGTGCGCGAGAGGTGATTTGA	89222
Qy	4811	CGGCATGATCATCTTCACCGGTGACGTACCGGCGCGGACGACAGTTCACGCGCGACGCT	4870
Db	89231	GGGCATGATCATCTGTCCGCGGTGATGCGCCAGAGGCTTTGAGACCATGTTTCAGCGCGCGGT	89172
Qy	4871	GGCGTTGACAGGTGTGCGCATGACAGTCCGATCCCGACACAGGTCCGCGACCCCGGCTTCGA	4930
Db	89171	GGCGTTGACAGGTGTCCACAGGGGTACCCACCCCGCGCATGTCCGCGCATGTGCGCTTCAG	89112
Qy	4931	CTCCCGACACAGGGGCGCGCGTGTGTGTGAGCGCATGTTGTTTGAAGGCCCATCTCCAGCGG	4990
Db	89111	TTCCGTGCGCACAGCGGGCGCGCGTGTGTGTGACCTCATCTGTTGTCCAGGCGCCATTCAGCGCG	89052
Qy	4991	GGCCAGAGAACCGCTGCGCGGTCCGATCTGTCCGCGCGCGCCACATGACAGGGGGTGCAGAA	5050
Db	89051	GGCCAGAGAACCGCGCGCGGTCCGATCTGTCCGCGCGCGCCACATGACAGGGGGTGCAGAA	88992
Qy	5051	CGCTTCGAGAACCGCGCGAATTCGCGAGATCGGTTCGATCGCGCTTCACGCCGTGCGCGG	5110
Db	88991	CGCTTCGAGAACCCCCGAGAACCGCGAGATCGGTGTGAGGCGCGCTTCATGGAATCACCC-C	88933
Qy	5111	TTGTGACCGCGCGACCGCGCAGCGGACAGAGCTTCGACGTGATGTTGACGTATGATGCTG	5170
Db	88932	CAGAGACCGCGCAGCTGACCGGGCCCCGATGAAGCGGTGCGCGCAGGTCAAGTTCAGT	88873
Qy	5171	TGCGCGAGCATTTGCGTGTCTGTGGCCGAGAGTATCATCACCGGAGTCCGGGGGACACCGCG	5230
Db	88872	GCGGAACCGCGCAGTCCGACAGTCCGGGAGCGCTTCAGAGAGGCGCTCGAAGGTGATTTCC	88813
Qy	5231	TGCTGCAAGTCCCGGGGACCTTCGACGAGCATGTACCGTTCCTGTTCGTTGCGGTGAAACCGG	5290
Db	88812	GCTTCCCGCGGTGACGCGCGCGCCACGCGAGTATGATGATGCGCGTGTGCGAAGGTGAGG	88753
Qy	5291	CCACCTTCTTCGAGTGTCCAGGCGGTGTACCGGATCTGGGTACCGGACGTTCCAGCAGG	5350
Db	88752	TGGCGCGCGGTGTTCGCGTGTGACCCCGAAGACATGCGGGGTCTTCGAAACCGCGCGGATTC	88693
Qy	5351	TAGTCGAGGTAGGTCCGGGCGGTGTTCTTCGGGACAGTCCGTTGATGTTTCCCGCTGACAG	5410
Db	88692	CGGTTGGCTGCGCGCAGAGTTGATGAGACAGGTCTCTGCGCGTGCATGTGACACCCCGCGG	88633
Qy	5411	TGCAACCGTCCCGCCCATCTCGGCGAGCTCTGTGCCACCCGCTCCGTTGCGCAGGTGACG	5470
Db	88632	ATCTGACAGGGCTTCAGAGGTGTCCCGGTCTGTATCTCATAGGTGTCCATGATGCGCAGC	88573
Qy	5471	AGCGCGTGTAGAGGTGCGCGTGTATCTCTTGTGACGAGAGGCGAGCATGCGGTGTGGG	5530
Db	88572	ATCTCTTCCACCGCGCGGCCACACAGCGCGCGTCCGCGCGGAGCGCGCGCGAGCTCCCGG	88513
Qy	5531	GGGAGAGACGCGGTTGATTCAGAGACCTCACCTCGCGGTGATCTGTGCTGTCACGAGAGG	5590
Db	88512	GGGTGGGTCAAGACAGAGAGGTGCGCGGTCCGAGCAGCTTACCGTGGTCTGTATCCCG	88453
Qy	5591	CCGAGAGTGTGATGATCTTGTCTCTGTGTTTCGATGCGCGGTGTCCGCGCGAGATCAC	5650
Db	88452	GCGACGAGCAGCAGAAACACCGTTCAGATGATGTTGTTCTTCGGAAGCGCGGTCCGCAAG	88393
Qy	5651	CCGTCGCGGTGATGTCCGCAACCGCGGGGTGCTTGGACGAACTCACGTGTGTCGT	5710

Db	88392	TCCGGCAGGTCGCACCACTGCCACACAGAGGTCTCTTCCCGGTTTGTCGCGCCGCGCGCG	88333
Oy	5711	ACGTCGGAATCCAGCTCATCAGAGGGTGTTCATTGTTGACACCGGGGTGCCGCCGCCGA	5768
Db	88392	ATCAGGAGGCGGACGCACTCGAACACAGGCTGCGGTATACGACACGCGCTTGGCGTCCGCGA	88275
RESULT 10			
LOCUS	AF521878	17665 bp DNA linear BCT 29-JUL-2002	
DEFINITION	Streptomyces narbonensis desosamine biosynthetic gene cluster.		
ACCESSION	AF521878		
VERSION	AF521878.1	GI:22001230	
SOURCE	Streptomyces narbonensis		
ORGANISM	Streptomyces narbonensis Bacteria; Actinobacteria; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 17665)		
AUTHORS	Bate,N. and Cundliffe,E.		
TITLE	The desosamine biosynthetic cluster of Streptomyces narbonensis, producer of nardomycin		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 17665)		
AUTHORS	Bate,N. and Cundliffe,E.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JUN-2002) Biochemistry, Leicester University, University Road, Leicester LE1 7RH, UK		
FEATURES	Location/Qualifiers		
source	1..1765	/organism="Streptomyces narbonensis"	
	/mol_type="genomic DNA"		
	/db_xref="taxon:67333"		
	<1..528	/gene="ndmA"	
	<1..528	/gene="ndmB"	
	/note="terminal thioesterase I (TEI) domain of type I polyketide synthase"		
	/codon_start=1		
	/transl_table=11		
	/product="NDMA"		
	/protein_id="AA088350.1"		
	/db_xref="GI:22001231"		
	/translation="AYLRPAGAPVTVLGHSGALLAEHLARFLBRAGRTPAQIVY DYPFHQEPLEIWSKQLGBGLFAGELBPMSPARTLLMGRYARFLAGRPERSSAPVL LVNASRLRGLDWPEERGDMRAWDLPHSVADVPGDHFTMRBDHAFAVAVALAMGLE RDYLRAGRTRACGADO"		
	525..1370		
	525..1370	/gene="ndmB"	
	525..1370	/gene="ndmB"	
	/note="thioesterase II (TEII)"		
	/codon_start=1		
	/transl_table=11		
	/product="NdmB"		
	/protein_id="AA088351.1"		
	/db_xref="GI:22001232"		
	/translation="MTDRLVANDSSLITRRFPHPNSAVRI.VCLPHAGGSAYFFRS BEIHPSVEALISVOYQRQDRAPETCLSEVBELAEHVATAEAMREGLAFPGSLGA SAVAEARLTLEORHGVRPEGGLVVGSRAPSLAPRLVQLDRAFLAIRIROLSTDSR PLDDIEIMELVT.PALRSDYKAETYEHRPSAKACPWLACGDPDPRKIPBAVEMER HTSGPCLAISYGSHPIYLDNDWHGICNDISHLVTRGAPDSPRYOPTRLIBGAARK WNQPR."		
	1367..2713		
	1367..2713	/gene="ndmC"	
	1367..2713	/gene="ndmC"	
	/note="putative NDP-deoxyhexose 3,4-isomerase"		
	/codon_start=1		
	/transl_table=11		
	/product="NdmC"		
gene			
CDS			

gene
CDS
2782. . 4062
/gene="nbmd"
/note="desosaminyl transferase"
/gene="nbmd"
/note="desosaminyl transferase"
/codon_start=1
/transl_table=11
/product="nbmd"
/protein_id="AAM89353.1"
/db_xref="GI:22001234"
/translation="MRVLTTSFAHHTHYGLVPLAMALLAAGHEVRVASOPALTDTIT
GGGLAAPVGTDLIHERYVMAGSPRENHPIAFDEARSFPLDMIDHALIDILAIAP
FVLANNMSMDIDIVGFARSMQOPDLVMEPTTGAVAAYOTGAARVAVGWGDS
ARKKVALRDROPPEHRSDPTLEHITWLTLDRFGASFREELITGOFTYDPPSLRLDT
GLPTGVQVQVLPNTGSVVPDWLAEPPASERVCLITGVASREVLGDVSGADILBALA
DIDIEVALTLAGORAEIRSYFKRTFTDFPMHALLPSCSAIVHNGAGCTVTAAVIN
APQVWLELMIDAPVKARAVALVBOGAGFPLPAGCTLPQAVRDAVVRILIDDSVILAAHR
LBEETHRDPTPAGIVPELBERLAAQRREPADARR"
4151. 4864
/gene="nbmg"
/gene="nbmg"
4151. 4864
/note="N,N-methyltransferase"
/codon_start=1
/transl_table=11
/product="nbmg"
/protein_id="AAM89354.1"
/db_xref="GI:22001235"
/translation="MYEVDHADVDLFTYLGKGDYAAASDIDADLVETRTPEASSLID
VAGCTGTLHEHTKEFGDTAGLESEDMTLFARRLPATALHOGMDPFLRGSRFSAY
VSMFSGVGLRTTALDLAAVASFAHLRPGVVVVEPWPPEFADGWASDVVRDGG
RTYARVSHSVRGDATHREHVTVAADPQGRVHRPSDHLITLPHRAYEAAFTPAAGLR
VELBEGSPSGFLGVGVA"
4308. . 7424
/gene="nbmp"
4308. . 7424
/gene="nbmp"
/note="beta glucosidase"
/codon_start=1
/transl_table=11
/product="nbmp"
/protein_id="AAM8935.1"
/db_xref="GI:22001236"
/translation="MHQOREKSYTYGKTRIPRVRSRSTTPRAITLAVVGTILAGTV
AAAPAPAGTIRVQYTSKAAELVAOMQTLDEKISFVHMALDDPDQNVGLPEVPLGIIP
ELAAAPAGTIRVQYTSKAAELVAOMQTLDEKISFVHMALDDPDQNVGLPEVPLGIIP
VNNITVHPGKRGYETFSBDPLVSSRTAAVAKIGIQAGKMTTAKHFPANQENRRFS
AMATVUBOTLRRIEIRPAPFASASKAGASPMCAVGNVGKPGSCNDELNNVLTITROMGF
OGVMSDMLATRGTDALITGDLQBEKVCYELPDGIPGSRPAKTPGCAIKOAVLNGTV
LEPAVGSASERIVQMDKFGLLATTPARPERDCKAQAVSRKAENGAVILIRYEQGA
LPLAAGSAGSINVTGPTADVPTGSGAHVVDSSAAAPLDITAKAPAGAGTVTYEQGA
EEFFRGIRPAAQGSIPAFNOGQOLEBPKAGALDTGLTVPADGEYRI SVKAGYATVQV
IGSHTIEAGQVGVKVSPLKLTGKTHKLTISGFAMASATPLSELGWTPEADATATLA
KYESARSKRTAVLVAFVADGTGVDRPNSLTPTGDDKLSANVADANPVTIVLVNAGSS
VLMPTSKRTAVLDMYPCQAGAEATALLVYDNPVPSKLTQSPALANQHAVAGDGRS
RYGVGNQOQTSBGIHVGYRMPDKENKVLFPFGHGLSTSTSGASAPTVATSTGGLK
RLQYTVNSGQRAQGVQVQVAYLQASPKVTAPOAEKGLVGTATVLAAGSKTYTVANVDR
VQYQYVDAASDRSTGSRSLQTCSSADLKNSSQVTVW"
complement (7514. . 8653)
/gene="nbmg"
complement (7514. . 8653)
/gene="nbmg"
/note="aminotransferase"

```

/codon_start=1
/cran1_table=11
/product="Nbmj"
/protein_id="AA08356.1"
/db_xref="GI:22001237"
/translation="MTARTETGRVPEFLDKAAVEBELPETDAIARVLDSGRYLLGPE
LEAFASPAFYCETDHAVGNSGMDALATLRGAIIGDDEVI VPSHTYI ASMWLAVSA
TGATVPVPEPNDEHPETLDPILVEKAIIPTRTRALLPYHLXGHPADLADRLDNRGLH
IYEDAAQAHGARVRRRIAGASVAAPSFYGKNULCGCGDGAAYTGDPALAEELRLM.
NRYGSRQRYSHEKTKGNSRLDEILCAVIRILALHDSNNGRSTLAAEYLSGLALGPG
IALPYTAGTDPVVMILFTVRTRDRLELDLHDSRQVDLTLYPVPVHLSPAYAQAPP
BSLPRASFAQVSMFLPIGHLERTQALRVIDAVREWAQAGR"
complement(8650. .9684)
/gene="nbmj"
/complement(8650. .9684)
/gene="ndmh"
/ncore="NDP-glucose 4,6 dehydratase"
/codon_start=1
/cran1_table=11
/product="Nbmj"
/protein_id="AA08357.1"
/db_xref="GI:22001238"
/translation="MRLVTGGAGFGTGHFVROLAGAYPDVAGEVIVLDSLTYYGN
RANLPEADPELRFPVHGIADADLAEELGVDAVVFPAESHDVRIIADASYFTGT
VNIQGTQILGCLADAGVGVVYSDYVGGISDGSNMSPLEBNSYKASGSD.
VARAHTRTGLDYLRTTRCCNNYGFQYHBEKILPLEVTMLDGCTLPIYGDAANREW
HTDDHCRCIALVALVAGARGVEVHIGGEBLTKRELITGILLBSLGDMSYRVADROG
HDLRRSLDGKILREILGYRPOIAFDAGLARTIRMYENRDMWEPLKATAPQPPATAQ
PATAVAGSA"
complement(9759. .10637)
/gene="nbmj"
/complement(9759. .10637)
/gene="ndmi"
/ncore="NDP-glucose synthase"
/codon_start=1
/cran1_table=11
/product="Nbmj"
/protein_id="AA08358.1"
/db_xref="GI:22001239"
/translation="MKGIYLAGSGGTRRLPATSVISKOILPYTNKPMIYYPLSLVLMG
GIREIQTISTPHIELFOSLLAGSGRHGILELDAVOKSPAGIADLIVGAGHIGDDTC
AIIIGNITPHAGQLVSLRDSIARLDGCVLFGYPVKDEBYGVABDVALGTGLGVK
PEKPRSHLAVNGLYLYDNDVNDVIAKNIRSPGKEITDYNVRVYLERGAELVLDVGRG
FALDITGHDSLIRPAQYQVYLEERQWIALEBIAFRMGITDAEACHSLQEGSRT
EYGSYMEIAGREGAP"
complement(10840. .12297)
/gene="nbmj"
/complement(10840. .12297)
/gene="nbmj"
/ncore="putative NDP-deoxyhexose 3,4-reductase"
/codon_start=1
/cran1_table=11
/product="Nbmj"
/protein_id="AA08359.1"
/db_xref="GI:22001240"
/translation="MTAPAPVATAPSAFCANHPGADLGAAYRVAVGSLAAGI VPSDEA
GSAIRHIVGLAVAYGNSPTPLSEAHHDIGVDRDAFRRLALFPEVPELRAVBSGPA
GAWQNTLPLLERRGVFSDALAGKVPFYSVGLYDQPTCMFRCHF CVRTAYRYPBSA
LDNGNMFPSVIDEIPAGNPHALYFSGSGLEPLNPGITLAAHALEHGLRPVYTN9F
ALTEPLTEPOGLMGILHAVRTSLYGSNDEYBQTKKAFARVABENLIRFOOLAEER
EAPINAGPAYIYIPGAAHLLDDVPIIADDSGGQRTDIPNIRIBDSGREGIDGLPO
EEBALEGEALAAFEERVRSRITPEGLHIDYIALNSLCAQADAEILIKIPATRYAPTHPO
VAOVLDLDGVYLYREAGPPDLGATRYIAGRVGPEPSTLTVRVYDVERGVAAARDG

```


Db 10903 GCGGCGCGTGAAGACTGTGTCGAAAGCCGTCATGAAATCTCGTCGCGCGGCGCG 10962
Qy 2515 GATGCTGCGGCGCGAAGCTGACGAACTGTGAGACGACGAGTGCAGGCTCGCGGT 2574
Db 10963 CACCGGCGCGCGCTTTCGAGTGTGCGGACGACCTCGGTGAGCGAGGTGCGGCGCC 11022
Qy 2575 GACCGGCGCGGCGAGTGTGCGGAGTGCAGCGCGCGAGCGCGGAGAAACCGGCTCGCGTA 2634
Db 11023 CAGCGGCGCGCGATGTACCGGCTGCGCGCTCGAGGTCCGAGAGCGCGGCTTCGCGGTA 11082
Qy 2635 CAGGTAGACGTCTTCGAGAGGTGTGATCTGACCGGCACTTCGCGATGCGGAGTCCGAGC 2694
Db 11083 CAGGTACAGTCCGCGAGAGGTGACCTGACCGGCACTTCGCGGAGTGCAGGAGCGG 11142
Qy 2695 CATGTTTCCGCGCGGATTCGCGCAACACTGCGGCTGACACCGCGAGCGAGCTCTCA 2754
Db 11143 CATGTGTGCGGCTTGTATCCGCGAGAGTTCCGCGTCCGCGCGCTTCAGGCGTTCAG 11202
Qy 2755 CGCGTAACCGAGTCCGCTGTGACATGCGCGGAGTCCGCTCGGCGGCGTATGTCAAGAAC 2814
Db 11203 GCGGTAGCCGTATGTGTGATGTGAGCCCGGCGGTGCGCTCGCGAGCCGCTCTCGAAGCC 11262
Qy 2815 GCGGAGGCGCTTCGCGCACTTCGCTCGCTTCGCGCACTTCGCGTCTGTCGCGAC 2874
Db 11263 GCGAGGCGCTCTCGAGAGTCCGCGCGCTCTCTCGGAGCGCTTCGCTTCGCGGCG 11322
Qy 2875 GCTGTAGTCTTCGCGAGCGGTGACGAACTGAGCGCGCGGTGCGCGGCTGACCTGTTGAG 2934
Db 11323 GCTGTATCTCTCGCGAGTGTGTGACGAGTGTGATGTGCTCTGCTCTGATCTGTGAG 11382
Qy 2935 TTGCGCATGATGATGACGACGATGATGAGCGGAGTGCAGCGCGCGCGCGAGATGATG 2994
Db 11383 GTGCGGATATGATGATGACGAGTGTGACGAGCGGTGAGCGCGCGCGCGAGACGAGTGA 11442
Qy 2995 GTTGAAGCGCGAGCGGAGCGGAGTGCAGCGCTTCGCGCGCGATTCGCGAGAGTCCCGAG 3054
Db 11443 GCGGAGGCGCGAGGTGTATCGCGCGCTTCGCGCGCGAGCGTGTGAGAGCGCGCGAG 11502
Qy 3055 GTTCTTCTGATCCCGTTCGAGAGCGCGAGCTTCGCGGTGTGAGCGCTGTGATCTGCTG 3114
Db 11503 GTTCTCGCGAGCGCGCGAGAGCGCGCTTCGCGGTGTGATCTGATCTGCTGCTC 11562
Qy 3115 GTTCAAGCGCGTACGAGGAGTGTGAGCGCGCGAGGTCGCGAGCGCGCGCGCGCGCG 3174
Db 11563 GTTGAAGCGCTTGAAGCGAGTGTGAGCGCGGTGAGCGCGCGCGCGCGCGCGCGCTC 11622
Qy 3175 CAGCGTCTGTTGCGGTGAGCGCGAGCGGTGTGTGAGAGCGGTGAGCGCGCGCGCGCG 3234
Db 11623 CAGGCTCGCTCGGTGAGCGCGAGAGAGTGTGTGTGAGCGGTGAGCGCGCGCGCGCTC 11682
Qy 3235 GCGCGCGTGCAGACCAACGCGCGCGATCCCGGAGTGTGACGCGTTTCAGGCGCGCGGA 3294
Db 11683 GCGCGCGTGCAGCG 11742
Qy 3295 GAGGTACATGCGCGGTGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3354
Db 11743 GAGGTACAGCGCGGTGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11802
Qy 3355 GTTTCG 3414
Db 11803 GTTTCGCGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11862
Qy 3415 ACAGCGGACATGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3474
Db 11863 GCGCGCGGACATGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11922
Qy 3475 CCGCGCGTACCG 3534
Db 11923 GCGCGCGAGCGCGCGAGTGTGAAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11982
Qy 3535 GTACTTTCG 3594

Db 11983 GTACGCCCG 12042
Qy 3595 GTTGAAGAGTCCCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3654
Db 12043 GCGGAGGAGCGCGCGGAGGCGGTCCGCGTGAAGCGCGCGCGCGCGCGCGCGCG 12102
Qy 3655 CCGGAGTGAACGCGGAGGCGCGCGTGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 3714
Db 12103 CCGGAGTGAAGGCGCGGTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12162
Qy 3715 TTGCGCGGTGTGAGTGTGCGGACCGGTGAGCGCGGTGACCTCGCGCTGAGCGCGCAC 3774
Db 12163 TCGCGCTGTGTGCGGACCGGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12222
Qy 3775 CAGGCGCGCGCGGTTCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3834
Db 12223 CCGCGCGCGCGGATCGCGCTTCGCGGAGTGTGCGAGCGCGCGCGCGCGCGCG 12282
Qy 3835 CGCGGTAGAGTGTGTTCATTCGACGCGGAGACCTTCGCGCGCTGTGTGCGAGGATC 3894
Db 12283 GCGCGGCGCGGTATCGGAGCGCTTCATGTGTGCGCGCGCGCGCGCGCGCGCGCG 12342
Qy 3895 ACGAACCGTTCGCTTCGCGTGTCCACTCGCGTGTATCAACGACCGCGTGTGTG 3954
Db 12343 GCGAGCGCGCGCGGTTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12397
Qy 3955 CAGTGTGATGACGTGTGCGACCGCGCGCGGTGTCTCACTGACACCGAGCGCGCGGT 4014
Db 12398 AGAGTTCAGCAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12457
Qy 4015 GAGGCGGAGCACCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4074
Db 12458 GAGGCGGAGCACCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12517
Qy 4075 TCGGATGAGCGGTATCTGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4134
Db 12518 CCGGATACCGGTCTCACTGTGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 12577
Qy 4135 CCGCGCGGAGATCGCGTGTGAGCGCGGTCAACGCGTGTGATGCGCGGTGCGTCA 4194
Db 12578 CCGCGCTTCAGGACCTTCATCAACGAGTGTGCGGTGTGAGCGCGGTGCGTCA 12637
Qy 4195 GAGTATCACTGATGTGTGTGTGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 4254
Db 12638 CAGGATACGATCTGTGTGTGTGTGACCGCGGTGTGCGCGCGCGCGCGCGCG 12697
Qy 4255 CCGGTACGTTCGCGCGACTCGGTGTGTGAGCGCGGTGTGTGACCGCGGTGTGTG 4314
Db 12698 CCGGAGAGTTCGCGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12757
Qy 4315 TCTCGCGGAGCGCGTGTGAGGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 4374
Db 12758 CCGCGGAGGAGCGCGTGTGAGGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 12817
Qy 4375 CCGTGTGTGCGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4434
Db 12818 CCGTGTGTGCG 12877
Qy 4435 GATTCGCTTCGCGCGAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4494
Db 12878 GATTCGCGCGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12937
Qy 4495 CCGCGT 4554
Db 12938 CCGCGT 12997
Qy 4555 CACCGCGT 4614
Db 12998 CCGT 13057
Qy 4615 CCGCGATCTTTCG 4674
Db 13058 CCGCGCGATCTTTCG 13117

Qy 4675 TCAGCGCGGCTCTGAGGGGATGACGAGCGGCGACGTGATCCGGGTGACACGACCGGTCG 4734
Db 13118 TGGCCGAGGTCGCGGGGCGTGAAGCGCGCGCCACCGCGTCCGGGTCCAGAGGTGCGGTGT 13177
Qy 4735 CCGGGTCGACGTGCGAGAACACCGGTGTGAGTCCGAGCCAGCTGCGCCGCTGCGCGGTGG 4794
Db 13178 CCGGGTCGATGTCGCGCAAGACCGGGGTGAGGCGGATCCAGCGCATGCGCGTGCGGGGTGG 13237
Qy 4795 CCGGAGAGTCATGACGGGATGATCACTTCAACCGGTGACGTACCGGCGCGGACGACCA 4854
Db 13338 CCGGAGAGTCATGACGGGATGATCACTTCAACCGGTGACGTACCGGCGCGGACGACCA 13397
Qy 4855 GTTCAGAGGCGACCGGTGCGGCTTGGAGGTGCGATGATGACCGGTACCCCGACGAGTCCG 4914
Db 13298 GCTGGAAGCCCGCGCGTGGCGCTTGAAGGTGCGCAAGCGGTGCTTGAACGCGCGGCGAGCGCG 13357
Qy 4915 CGACCCGCGGCTCGAATCTCCGCAACGAGGGCGCGCGGTGTGATGACCAAGTGTGTTCA 4974
Db 13358 CGACCGCGCTCTCGAATCTCCGCAACGAGGGCGCGCGGTGTGATGACCAAGTGTGTTCA 13417
Qy 4975 GGGCCCACTCGAGCGCGGCGGAGAACCGCTGCGCGGTGCGCGATGCTGCGCGCGCGCGCT 5034
Db 13418 GCGCGCGGTGAGCGCGCTGTAAGCTTGGCGCGGTGATGCGGTGCGCGCGCGCGCGCA 13477
Qy 5035 GCAGCGGGGTGAGGAGCGCTCGGAGACCGCGGAGATGCGGATGCGGTGCGGTGCGGT 5094
Db 13478 GGAACGCGCTGATCGAAGCGCGCGGTGCGCGGAGAGATGCGGATGCGGTGCGGT 13537
Qy 5095 TCACG 5099
Db 13538 TCACG 13542

RESULT 11
AF079762 12441 bp DNA linear BCT 27-OCT-1998
LOCUS AF079762
DEFINITION Streptomyces venezuelae desosamine biosynthetic gene cluster,
complete sequence.
ACCESSION AF079762
VERSION AF079762.1 GI:3789892
KEYWORDS
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Xue, Y., Zhao, L., Liu, H. W. and Sherman, D. H.
TITLE A gene cluster for macrolide antibiotic biosynthesis in
Streptomyces venezuelae: architecture of metabolic diversity
Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
JOURNAL PUBMED 9770448
2 (bases 1 to 12441)
REFERENCE Zhao, L., Liu, H. W. and Sherman, D. H.
AUTHORS Direct Submission
TITLE Submitted (22-JUL-1998) Department of Microbiology, University of
JOURNAL Minnesota, 420 Delaware Street SE # 1060, Minneapolis, MN 55455,
USA

FEATURES
SOURCE Location/Qualifiers
1..12441
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
119..1327
/gene="desVII"
119..1327
/gene="desVIII"
119..1327
/gene="desVIII"
/codon_start=1
/transl_table=1
/product="putative tautomerase"
/protein_id="AAC68676.1"

/db_xref="GI:3789893"
/translation="MTDDLTGALTOPPLGRTVRAVADRELGLHLETRGIIHWIHAANG
DPATYALRGADDDPYPAVERARALISFPTSGWTADALAAIISCTDFGSGAD
GVPPPOOVSYGBGCLEREOVPAAGDDVGGORAVBGIIHETLEGLAPDSASYA
FELGFPVPAATYAAAAYIGPADRADPADLLERLPRPISDLIAPQSIRTYAAG
ALBELTALDSDSDSPGALLSALGATVAUUTGNATIALAHBQWELCDRLPDLAA
AVETLRYDDPVOLDARVVRGTELRGRLPAGAHVVLTAATGRDPEVTPDERFL
ARPDAAHNLHPAGPYGVALSVRLQAEVALRTLAGRFGLRQAGDVLPRRPAVR
GPLSVPS89"
1475..2755
/gene="desVII"
1475..2755
/gene="desVII"
/codon_start=1
/transl_table=1
/product="glycosyl transferase"
/protein_id="AAC68677.1"
/db_xref="GI:3789894"
/translation="MRVLTSTFAHHTHYGLVPLANALLAAGHEVRA9QPALDTIT
GSGIAAPVGTDLIHEYRVRMGEPSPNHPALAFDARPEPLDMWHAIGIETIAY
FHLIANDSWDDLVDFAVSMODPLVMEPTTGAAGAAQVGAHARVMDGVMS
ARRKFLVLRDROPEHREDPTAEMLTTLDRYGASFEEELTQFTIDTPPSRLDT
GLPTVGRVYPVGTSTVPPWLSBPAPRVCULTIGVARSERVGGDVGSGDILFALA
DLIDELVATIDAGRAIRIRYPKTRTPTDVPVPHALLPSCSAIIHGGAGVYAAVN
AVQVMLAEIEMDAVVKARAAGEAGFLLPAELTPOAVSDAVVRLIDPSVATAAR
LRBETFDPTPAGIVPELRLAQRHPPADARH"
2845..3558
/gene="desVI"
2845..3558
/gene="desVI"
/codon_start=1
/transl_table=1
/product="N,N-dimethyltransferase"
/protein_id="AAC68678.1"
/db_xref="GI:3789895"
/translation="MYEVDADVDYDLYLRGKDYAAEADIDAVSRPPEASGLID
VACGTGHLHFTEFEDTAGLEISBMLTHAKRLPDATLHQDMKDFRIGKFSAY
VSMFSVGYGKTEBELGAASFAEHLDPGVVAVVEPWEPETPADSVASADVVRDG
RTVARVSHSVREGNATMEVHFTVADPGKVRPSDVHLLTFHQAEYEAFAAQLR
VEYLEGSGRGLFVGVA"
3625..6054
/gene="desR"
3625..6054
/gene="desR"
/note="putative extracellular protein involved in a
resistance mechanism"
/codon_start=1
/transl_table=1
/product="beta-glucosidase"
/protein_id="AAC68679.1"
/db_xref="GI:3789896"
/translation="MTGKTRIPRVRRKRTTPRAFLAVVGLTAGTTVAAPAAADT
ANVOYTSRAAEVVAQMLDEKISFVMAALDDPDONVGLPGVPLGIPELAAAGPNG
IRLVGTATLAPVALASTFDTDMAISYGMKRDGRALNDQVLPMMNITVPHG
GRNETHSEBDLVSSRTAAVAQIKGIAQGLMTAKHAPANNQENNRSGVNAVSDTL
REIFPAPFASSKGAASFMCAVYNGKSGCNDLNNVLRQWOGSVGMVMSDWLA
TPGDIATIKGLDQMGYELPDGVKPKSPAKFEGALITAVLNGVPEAAYVRSAS
RIVGDMERKGLIATPAAPRPERDKAGAAVSRKVAENGAVILRNQGLPLAAGAGKS
IAYIGPRAVDPKVTGSGSAHYVPPSAAPLDTITAPRAGAGATVYVTEGSETFQDIPA
GNLSPAFNQGHLSPKAGALYDGTITVPADGRIARAVRATGATYATQSGHTIBAG
VYGVSSPLKLLTGTHKLTISGFAMSAITPLISLGMWTPPAADATTAKAVESAKAR
TAVFAVDDGTEGVDPRPMLSLPGTODKLSAVNADANPTIVLVTGSSVLMPLMSTKR
AIVLDMWYDGAAGATATALLYGVDPGSKLQSFPALENQAVAGDPTSPYGVNQOT
YREGIHWYRPFDEKENVKPLFPFGHGLSYTSFQSGAPVAVVTSGGKLVITYVNSGK
RAGGVQATYAGASPNTYAPAKKLVGYTYVSLAAGAKTYTVANVDRQVOTG886A
DLRGSATVNV"
complement(6143..7282)
/gene="desV"
complement(6143..7282)
/gene="desV"
/codon_start=1
/transl_table=1
/product="transaminase"

gene
CDS

```
gene
CDS
    /protein_id="AAC68680.1"
    /db_xref="GI:3789897"
    /translation="MSSRAETPRVFLDLKAAYBELAETDAIARVLDGSRVILGPE
    LGEPRFAAYCETDHAIVGNSMDLQIALRGIGRGDEVIWPSSTYASLVAVA
    TGAIPVPERHEDEPTDPLVKAITPRRLGIPHLVGHFADMDLRELDHGLH
    IVEADAQHGAKTRGRIIGAGSSVAASFPGKNLGFSGGAAYTGDPELAEKRL
    RNSGRQKSHETKTRSLRDEMOAVLRILALHDSNGRSLAAEYVSGIAGLPG
    IGLPVAPDTPVWHLPTVTERDERLSHLDAAGIDTTLTHYVPVPLSPAYGEAP
    EGSLPRASFARQVLSLIGPHLERPOLAVITAVREMARVDA"
    complement(7279. .8292)
    /gene="desiv"
    complement(7279. .8292)
    /gene="desiv"
    /codon_start=1
    /transl_table=11
    /product="TDP-glucose-4,6-dehydratase"
    /protein_id="AAC68681.1"
    /db_xref="GI:3789898"
    /translation="MRLVTGAGFIGSHFRLQLAGAPDPADEVITVLSLYAGN
    RANLAPVDADPRLRFVGDIRDAGLARELGVDAIVHFAESHSDESISAGSVFET
    NVOGTOTLCAVDAGVGVVSTDEVSIDSQSWHTESSPLRPNSPVYASAGSGL
    VARAHRTVGLDVIRTRCCNNVGYQHPGLIPLPTVNLIDGGLPLVYGDANREY
    HTDHCKGLALVLAGGAEIYHIGGLILNRBELTGLLDSGADWSVRAVDKRG
    HDLRYSLDGEIERELGYRQVSPADGLARTVWYRENGMWBPLKATAPOLATAVE
    VSA"
    complement(8424. .9302)
    /gene="desiit"
    complement(8424. .9302)
    /gene="desiit"
    /codon_start=1
    /transl_table=11
    /product="alpha-D-glucose-1-phosphate
    thymidyltransferase"
    /protein_id="AAC68682.1"
    /db_xref="GI:3789899"
    /translation="MKGIYLAGSGTRLHPATSVISKQILPVYKEMTYPLSLMIG
    GIREIQIISTPOHILFQSLGNGRHGLELDYAVQKEPAGIADALLVGAHIGDTC
    ALILGNIIFHGPGLYTLRLDSIARLDGCVLFGVPDPERGVAVDAGRLDIYER
    PKYRNSNIAVTGLYLDNDVVDIAXNIRPSRGLERTDYNRYILRGRAELVNLRG
    FAWLDGTGHSLSLRMAQYOVVLEERGCVWLAGLEIAFRMGFIIDAECHGLGSLRT
    EYGSYLMEIAGRGAAP"
    complement(9495. .10952)
    /gene="desiit"
    complement(9495. .10952)
    /gene="desiit"
    /codon_start=1
    /transl_table=11
    /product="putative reductase"
    /protein_id="AAC68683.1"
    /db_xref="GI:3789900"
    /translation="WTRAPLSATAPARCAHPADLGAAVHAGTLLAAGLVPRBA
    GTARHVRILAVYGNSPPTPLEARHDLGVDADARRLALFGVPELRTAVETGA
    GAWKNITLLEQRGVFDALARKPVPSVGLYPPCTMFRCHFCVRVYGAETGSA
    LAGANMFRSVIDEIPAGNPSAMYFSGGLEPLTNPGLSLAAATDHGRPLVYTSF
    ALBERTLEROPGLMGALRTSLYGNDEYEQTKKAFFRVRNRLRFPOLRER
    ESPINLGPAYIVLPGASRLDLVDPIADLNDAGQRTIDFVYIREDYSGRDGKIPO
    ERBAEIQBALMAFEREYRERTPELHIDYGTALNSLRTGDAELRLIKPAMRTPARPO
    VAVQVLDLQDVVYLREAGFPDLDGATRYLAGRTVPTSLTEVVRDEVERGEVAADVQ
    DEYFMDGPDQVYVARLNOLERDADGMEBARGFIR"
    complement(10949. .12196)
    /gene="desiit"
    complement(10949. .12196)
    /gene="desiit"
    /codon_start=1
    /transl_table=11
    /product="4-dehydroase"
    /protein_id="AAC68684.1"
    /db_xref="GI:3789901"
    /translation="MKSALSDLAFFGGAAPDPLVGRPNRIDARLYRDLRALDS
    QWLSNGPLREFEERVAGLAGRAHVAATCAVAGIQLAAHAGLGEVIMPMTPAA
    TPLALWIGLTPVADIDPTGNLDPDVAAATPTSTAVGVLHMGRCADQDLAKY
    ADHGLRLYFDAAHALGACVDRGASLSDAEVFSHATKAVVAFEGGAVDDADLA
    ARIRALHNFGLDGGSPAGGTAAKMSAAAMGLSLDAFVIDIRNRNNAAYREH
```

ORIGIN

LADLPGVILVADHDHGLNNHQQVIVEIDEATYIHRDLVMEVLKAEGRVTRAFSPGC
HELEPYRGOPHAPLPHTBRLLAAARVLSLPSTALIGDDDIRVADLRLRCATRGHETLAR
HRTAPAPLAAAPOTSPTTIGRSR"

```
Query Match 14.6%; Score 1309.4; DB 1; Length 12441;
Best Local Similarity 67.7%; Pred. No. 6,3e-55;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

QY 2369 GGGCGGTCGACGCCGTTCAAGACGCGCGGATCAGGCCGATGCTGATCCACCGTCG 2428
    |||||
DB 9472 GGGGGGATCGGGCGGGGTGGGTGCGCGCGAGAAAGCGCGGCGCTCCCGACCGCTCC 9531

QY 2429 GCGACGTCCTCGTTCGAGTTGGTTAGCGCGGGGGGTACACGACTGTCGAAACCGTGAAG 2488
    |||||
DB 9532 GCGGCGTCGCGCTTCAGCTGCTTCAGCGCGGGCGGTGAACGCTGATCGAAAGCGTCCATG 9591

QY 2489 AAGAACTCGTCCCGCGCTGCGATGCGATGCTGCGCGCGCGCGCGTGAAGAGTGTGACG 2548
    |||||
DB 9592 AAGTACTCGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9651

QY 2549 ACCGAGTGAAGCTCGGTCGCGGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2608
    |||||
DB 9652 ACTTCGATGAGGAGGAGTTCGCGGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9711

QY 2609 AGCCCGGGGAAACCGCGCTCCCGGTACAGGTAGACGTCCTCCGACGAGTTCGACCG 2668
    |||||
DB 9712 AGGTTCGGGAAAGCGCGCTTCGCGGTACAGGTACAGTTCGCGCGGAGAGTTCGACCTGAC 9771

QY 2669 GCCACCTGCGGATGCGCGCGGTGCGCGCGCATGTTTCGCGCGCGATCCGCAACAGTGGCG 2728
    |||||
DB 9772 GCGACCTGCGGCGGTGCGCGCGGTGCGCGCGCATGTTTCGCGCGCGATCCGCAACAGTGGCG 9831

QY 2729 TCGACACCCCGACGACGCGCTTCGCAACGCGTAAACGAGTGGTTCGATGCCGCGGTC 2788
    |||||
DB 9832 TCGGCCCCGGGTGCGCGCGCTTCGCAACGCGGTAAACGAGTGGTTCGATGCCGCGGTCG 9891

QY 2789 CGCTCGCGCGCGTACGTCAGCAAGAACCGGCGGAGGCGCTCCCGCACTCGCTTCGCGCC 2848
    |||||
DB 9892 CGCTCGCGCGCGCTTCGCAAGAGGCGTTCGAGGCGCTTCGCGCGCGCGCGCGCGCTTC 9951

QY 2849 TCGGACAACTGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2908
    |||||
DB 9952 TCGGCGACGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10011

QY 2909 GCGCGGTGCGCGGTGACCTGTTGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2968
    |||||
DB 10012 GTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10071

QY 2969 TTGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3028
    |||||
DB 10072 GAGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10131

QY 3029 GCGCGCATCGCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3088
    |||||
DB 10132 GCGCGCATCGCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10191

QY 3089 CCGGTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3148
    |||||
DB 10192 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10251

QY 3149 AGGTCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3208
    |||||
DB 10252 AGGCGCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10311

QY 3209 TGAAGGTGCGAGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268
    |||||
DB 10312 TGAAGGTGCGAGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10371

QY 3269 TTGCTCAACGCTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3328
    |||||
DB 10372 TTGCTCAACGCGCTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10431
```

QY 3329 TCGTCATGACCGAGGTGAGAGATCTGTTGCGGAGCACACGACTCCGCGTCGTAATTG 3388
Db 10432 TCGTCATGACCGAGGAAATGAGCGTTGCGGCGTCGAGGCGGACGCGGTCGTAAGCG 10491
QY 3389 GCACCTGTACCCCTGATGCGAAGATGACAGCGGAAATGCACTGTCGCGCGCGGTTAAGC 3448
Db 10492 GCGCGGTGACACGAGACGCAAAATGCGAGGAAATGCAAGTCCGCGCGGTTAAGG 10551
QY 3449 CCGACGCTGTACGAGGAAAGCGGAGCTCCGCGCTACCGGAGCTTCAACGCGCGCGAGCG 3508
Db 10552 CCGACGCTGTACGAGGAAAGCGGAGCTTCCGCGAGCGCGCGCTGTGAAAGACCGCGCTGT 10611
QY 3509 TTACGCGCGCTGATCGTGTGCTCAGTACTTTCGCGCGGAGCGCTGTCCACCGCGTG 3568
Db 10612 TCGAGCGGAGACAGAGGTGTTCTTCAGTACGCCCGCGCGGAGCGGTCTGACCGCGGTG 10671
QY 3569 CCGAGCTGCGGAGATGCGATGCAACAGTTCAGCAGCTTCGCGAACCGCGCGCGGTCCAGG 3628
Db 10672 CCGAGCTGCGGAGACCTGCGCGCAAGGCGCGAGAGCGCGCGAAAGCGTCCCGGTGACG 10731
QY 3629 TCGAACGAGCGCGCATCTGCTCCAGCGGAGGTGAACCGGCGGAGCGCGTGTAGGCGGTG 3688
Db 10732 CCGAGCTGTCGCGCGGCTCTCTCCAGCGGAGGTGAAGGCGGCTGTGCTGACGACGCGG 10791
QY 3689 AGTTCGCGAGCGCGACCGCTGTCTGTTCGCGCGGTTCGATGCGAGCACCGGTGAGCGG 3748
Db 10792 AGCGGAGACAGGTGCGCGCGGCGGTTCGCGCTGTGCGCGGAGACGAGCGCGCGGCG 10851
QY 3749 GTGACCTGCGCGCTGAGCGCGCGCACCGCGCGCGGTGCGCGGTGCGCGCGG 3808
Db 10852 GCGAGGCTGTCGCGCGAGCGGCTGAGCACCGCGCGCGCATGCGCTCCCGGAGTGCAG 10911
QY 3809 TCCGCGATCTCGCTGCGTACGCGGCGGTGCGCGAGAGGTGCTTCATGACGTGCAAC 3868
Db 10912 CGTTCGCGCGGCGGCGGTGCGGAAAGCGGCGGCGGTGCTGCGAGGCTTCATGCGG 10971
QY 3869 CTTTCGCGGTGTGTGTCGAGAGATCAAGAACCGTTGCGTTCGCTGCTCCACTCGC 3928
Db 10972 GCGTGTGATGTCTGCGGCGGCGCGGAGCGGCGGCGGTGTCGCGGTGCGGCGGCG 11026
QY 3929 GTTGTATCAAGCAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3988
Db 11027 GGTCACTTTCGCGCGCGGTGTCGAGACGCGAGAGGTGCGAGACCGCGCGGTATTC 11086
QY 3989 CTCACGTGACACCGAGGCGCGGTGCGGAGCGGAGCACCGGTTGCGAGCGCACTCGGT 4048
Db 11087 GTCTGTGCGGATGCGGCGGTGCGGTCGAGGAGACGACGCGCGCGGAGCGGTTCGT 11146
QY 4049 CTGTGTCAAGCGCGACGCGGTGCTCGTGTGCGGTGAGGCGGTCACTGTGTGTGTGTGT 4108
Db 11147 GTGCGGACGAGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCG 11206
QY 4109 GAAGTAGGCGCTGT 4168
Db 11207 GAAGTAGGCGCGGT 11266
QY 4169 GATGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTG 4228
Db 11267 GATGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTG 11326
QY 4229 GCGCGGAGCGAGTGTGACGCGGTGACGCGGTGACGCGGTGACGCGGTGACGCGGTG 4288
Db 11327 GCGGT 11386
QY 4289 GTGTGTGACCGCGGT 4348
Db 11387 GTGTGTGCGCGGTTCGCGGTGATGACCTGCGGAAACCGGTGAGAGGTGTGAGCGCAT 11446
QY 4349 GCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4408
Db 11447 GCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11506
QY 4409 GCCGATGCGGAGT 4468

Db 11507 GTGAGACCGGAGT 11566
QY 4469 CACCGCGCGCGCTCGAAGCGGTGACCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4528
Db 11567 GACGCGCGCGCTCGAAGCGGTGACCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11626
QY 4529 GTGACCGGATCCGCGCACCGGACGTCACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4588
Db 11627 GTGCGCGAGGTGCG 11686
QY 4589 GAAAGAGAGCTTGAAGCTGT 4648
Db 11687 GAAGTACAGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11746
QY 4649 TCGGCGCGCACAGGT 4708
Db 11747 GCG 11806
QY 4709 GTGTGCGGAGTGTGACCGAGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 4768
Db 11807 CTGTGCGGAGTGTGACCGAGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 11866
QY 4769 GAGCGAGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 4828
Db 11867 GATCAGCGAGTGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 11926
QY 4829 GTGTGCGTGTGCG 4888
Db 11927 GTGTGCG 11986
QY 4889 GCACTGTGCGTGTGCG 4948
Db 11987 GCGATGTGCG 12046
QY 4949 GCGGTGTGTGAGCGCGGT 5008
Db 12047 GCGGTGTGTGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12106
QY 5009 GTGCGCGATGTGCG 5068
Db 12107 GTGATGTGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12166
QY 5069 GATGCGAGATGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5099
Db 12167 GATGCGAGATGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 12197

RESULT 12
BD217172 12441 bp DNA linear PAT 17-JUL-2003
LOCUS BD217172
DEFINITION DNA encoding methymycin and pikromycin.
ACCESSION BD217172.1 GI:33026942
VERSION BD217172.1 GI:33026942
KEYWORDS JP 2002536959-A/2.
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 12441)
SHERMAN,D.H., LIU,H.W., XUE,Y. and ZHAO,L.
DNA encoding methymycin and pikromycin
Patent: JP 2002536959-A 2 05-NOV-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA
OS Streptomyces venezuelae
PN JP 2002536959-A/2
PD 05-NOV-2002
PF 25-JUN-1999 JP 2000557373
PR 26-JUN-1998 US 09/105537
PI DAVID H SHERMAN, HUNG MEN LIU, YONGQUAN XUE, LISHAN ZHAO PC
C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12P17/ 42, C12P19/62,
PC C12N15/00, C12N5/00

CC DNA encoding methymycin and pikromycin
FH Key Location/Qualifiers
FT source 1..12441
PT /organism='Streptomyces venezuelae'
Location/Qualifiers 1..12441
/organism='Streptomyces venezuelae'
/mol_type='genomic DNA'
/db_xref='taxon:54571'

Query Match 14.6%; Score 1309.4; DB 6; Length 12441;
Best Local Similarity 67.7%; Pred. No. 6.3e-85;
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

QY 2369 GGGGGGGTGCACCCGCTTCAGACGCGCGGATGAGGCGCCGATGCTGATCCACCGCTCG 2428
DB 9472 GGGGGGATCGGGGCGGGGTGGGGTCAAGCGAGAGAGCGCGGCTTCCTCCAGCCGCTCC 9531
QY 2429 GCGACGTCCTCGTTGAGTTGTTGAGCCGGCGGTCACCGACTGCTGATGAAACCGTGCAGG 2488
DB 9532 GCGGCGTTCGGCTTCAGCTGCTTCAGCGCGGCGGTGACAGCTGATGAAAGCCGTCCATG 9591
QY 2489 AAGAACTGTCCTCCGCGCTGCGGATCGATGCTGCGGCGCGACGTTGACGAAGTGTGACG 2548
DB 9592 AAGTACTCGTCGCGCTGACGCGCGGCACTCGCGCGCGCTGACGAAAGTCCCTGACG 9651
QY 2549 ACCGAGTGCAGGCTCGGCTCGGCGGTGACCCGCGCGGCGACGTTAGCGGGTGGCCCGCC 2608
DB 9652 ACTTCGATGAGGAGGTGTCGGGGGTGACCGCGCGCGCGATGATGAGCGGGTCCGCTCC 9711
QY 2609 AGCCCGGGGAAACCGGCTTCGCGGTACAGGTAGACGTTCCGACGAGTGTGATCTGACG 2668
DB 9712 AGGTCCGGGAGAGCGGCTTCGCGGTACAGGTACAGTCCGCGAGAGATGACCTGACG 9771
QY 2669 GCCACTGCGGATGAGCGGCTCGGCGGATGTTTCGCGCGGATTCGCAACAGCTGCGCG 2728
DB 9772 GCGACTGCGGGTGGCGGGTGGCGCGGATGTTGAGCGGGCTTGAATCGGACAGATTGCGG 9831
QY 2729 TCGAACACCCGAGAGCGGCTTCGAACGCGTTAACCGAGTGGTTCGATGCTCCGCGG 2788
DB 9832 TCGGCCCCGAGTGGCGAGGCTGTTCAAGCGCTGACCGTGAATGATGAGTCCGCGGGT 9891
QY 2789 CGCTCGCGCGGCTGATGACGAGACCGGCGGCGGCGCTCCGCGACCTCGGCTTCGCGCC 2848
DB 9892 CGCTCCGCGAGCCGCTTCCTGAAAGCGTTGAGGGGCTCTCGAGCTCGCGCTCTCC 9951
QY 2849 TCGAACAACTGCGGCTTCGCGCGGCTGATGCTTCGCGCGAGCGGTTGACGAAGTGCAGC 2908
DB 9952 TCGGAGAGCTGCGGCTTCGCGCGGCTGATGCTTCGCGGAATGTTGACGAAGTGCATC 10011
QY 2909 GCGCGGCTGCGGCTGATGCTTCGTTGAGTTCGCGCGATGATGACCAAGTGCATGAGCGG 2968
DB 10012 GTCTTCGCGCTGCGCGGCTGCTGTTGAGGTGCGCGATGATGACCAAGTGCATGAGCGG 10071
QY 2969 TTGGCGCGCGCGCGGCGAGATGATGTTGAGCGCGAGCGCGGCTGCGCGGCTTCGCG 3028
DB 10072 GAGGCAACGCGCGCGGAGCAAGATGAGGCGAAGCCAGGTTGATGCGCGCACTGCGCTCG 10131
QY 3029 GCGCGCATCCGAGAGATCCCGGAGGTTCTTCTGATCCGTTGCAAGGCGCGACGCTTG 3088
DB 10132 GCGCGGAGCTGCTGAGAGGCGGCGAGGTTCTCGCGGACCGCGCGGAGAGCGCGCTTCG 10191
QY 3089 CGGCTGCTGCGCTGATGCTTCGTTGTTGATGCGCGGATGAGGAGGATGAGCGCGCGCG 3148
DB 10192 CGGCTGCTGCTGATGCTTCGTTGTTGAGGCGCGTGAAGCGAGTGGCGATGCGCTGC 10251
QY 3149 AGGTCCCAAGCGCGGCTGAGCGGCGGCGAGGCTGTTGCGTGAAGGCGGAGAGGCTTG 3208
DB 10252 AGGCCCCAGAGCGCGGCTGAGCGCTCAAGGATGCGCTCGGTGAGCGGAGAGGATTCG 10311
QY 3209 TAGACGCTGAGGCGGAGCGCGCGCGCGCGCGCTGCGACCAACCGCCGATGACCGCGG 3268

DB 10312 TAGACGCTGAGGCGCGAGCGCGCTGCTGCTGCGGTGCGCGGCGAGGCTCCCGAGGCGG 10371
QY 3269 TTGCTCAACGCTTCAGGCGCGCGGAGAGGATGATGCGCGTCCGCTTCCTCCAGCGCAC 3328
DB 10372 TTGCTGAGCGGCTTCAGGCGCGCGGAGAGGATGATGCGCGAGGCGGTTCGCGCGGTATC 10431
QY 3329 TCGTCAATGACCGAGGCTGAGGATCTGTTGCGCGGCGACCAACGATCTCGGCTGATCTTG 3388
DB 10432 TCGTCAATGACCGAGCGGATGATGCGGATGCGCGCTCGAGGCGGAGCGGCTGATCGG 10491
QY 3389 GCACTGTGACCCGTTGACGAGAGTGAACAGCGGAGATGACGCTCGGCGCGGCTGAGAGC 3448
DB 10492 GCGCGGCTGACAGAGCGGAGAGTGGACGCGGAGATGAGGATGAGGCTCGGCGCGGCTGAGG 10551
QY 3449 CCGACGCTGATCGGAGAGGCGGCTTCGCGCGTTCGCGAGCTTCGACGCGCGCGCGCG 3508
DB 10552 CCGACGCTGATCGGAGAGGCGGCTTCGCGCGTTCGCGAGCGCGCGCTGAGAGAGCGCGCTGT 10611
QY 3509 TTGAGCGGCGCTGATGCTGTTGCTCGATGCTTCGCGCGCGGCGCGCTTCGACGCGCTG 3568
DB 10612 TCGAGCGGAGAGCGGCTGCTTCGATGAGCGCGCGCGGCGCGCTTCGACGCGCGGTG 10671
QY 3569 CCGAGCTGCGGATGCGATGGAACAGGTGAGCAGCTCGCGCGAACCGCGCGCGCTGACAG 3628
DB 10672 CCGAGCTCGGAGAGCTGCGCGAACAGGCGGAGGCGCGGAGGCGCTTCGCGTGCAGC 10731
QY 3629 TCGAACCAACGCGGATGCTGCTTCGACGCGGATGAACGCGGAGCGCGCTGATGAGCGGTG 3688
DB 10732 CCGAGCTGCTGCGGCGCTTCCTCAGCGGCGGATGAAGGCGCTGTTGCTGATGACGAGCG 10791
QY 3689 AGTTGCGGAGAGCGGCGCGCTGCTGCTGCGGCGGTGCGGATGCGGCGCGCGCGCGGCGG 3748
DB 10792 AGCGGAGAGAGGTGCGGCGCGGTGCTTCGCGCTTCGCGCGCGCGCACAGAGCGCGCGG 10851
QY 3749 GTGACTCGGCGCTGAGCGCGCGCGCACAGCGCGCGCGGCTGCGGCTGCTGCGCGCGG 3808
DB 10852 GCGAGGCTGCGGCGG 10911
QY 3809 TCGCGATTCGCTGCGGTGCGCGCGGTGCGCGGATGAGGCTTCATGACGTCGCAAC 3868
DB 10912 CGTTGCGCGCGGCGGCTGCGGAGAGGCGGCGGCGGCTGATGAGGCGTCCAAATG 10971
QY 3869 CTTTCGCGCTGCTGCTGCGGAGATGACGAAACGTTGCGCTTCGCTTCGCACTCGCG 3928
DB 10972 GCGTGAATTCGCGGCG 11026
QY 3929 GTTGAATCAACGACGCGCTGCTGCTGCTGCGGATGAGTGAAGTGCACACCGCGCGATG 3988
DB 11027 GGTCACTGACCGG 11086
QY 3989 CTCACTGACACCGAGGCGCGGCTGCGGAGGCGGAGCGCGCGCTTCGCGCGAGCACTCG 4048
DB 11087 GTCTGCGCGGATGCGGCTGCGGCGGAGGAGCGAGCGCGCGCGCGCGCGCGCGG 11146
QY 4049 CTGCTGCGCGCGCGCGCGGCTGCGGCGGATGAGGCGGCTGCTGCTGCTGCGCGCGG 4108
DB 11147 GTGCGGCAACCGGCGCGGCTGCGGCGGATGAGGCTGCTGCTGCTGCGCGCGCGG 11206
QY 4109 GAAATGAGGCTGCTGCGGAGCACTTCGCGCGCGGAGATGCGCTGCGAGCGCGGCTGAG 4168
DB 11207 GAAATGAGGCGCGGCGGATGAGCGCGCTTCGCGCGGAGCTGCTGAGGAGGCTGCGG 11266
QY 4169 GATGCGGCTGCGGCGGCTGCGGCGGATGATGATGATGATGATGATGATGATGATGATG 4228
DB 11267 GATGCGGCTGCGGCGGCTGCGGCGGATGATGATGATGATGATGATGATGATGATGATG 11326
QY 4229 GGGCGGAGCGAGTGAACGCGCGGATGAGCGCGGATGAGCGCGGCTGCTGCTGCTG 4288
DB 11327 GCGGCTGCTGCTGCGGAGGAGCGCGGAGGCTGCGGAGGCTGCTGCGGAGGCGG 11386
QY 4289 GTGCTGACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4348
DB 11387 GTGCTGCGCGGCTTCGCGGATGATGATGATGATGATGATGATGATGATGATGATG 11446

[illegible]

AUTHORS
 TITLE
 JOURNAL
 Submitted (16-ANG-2002) Yojiro Anzai, Toho University, Faculty of
 Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba, 274-8510,
 Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel: 81-47-472-2072,
 Fax: 81-47-472-2086)
 FEATURES
 SOURCE
 location/Qualifiers
 1. 58343
 /organism="Mycrospora griseorubida"
 /mol_type="genomic DNA"
 /db_xref="taxon:28040"
 complement (<1. .910)
 /gene="mycF"
 complement (<1. .910)
 /gene="mycF"
 /codon_start=1
 /transl_table=1
 /product="methyltransferase"
 /protein_id="BACS7022.1"
 /db_xref="GI:28394149"
 /translation="MSRSTVELYLDLKRVTNSNFTYQDATVHAGLITAEAFYBARE
 SGEYFVATHIMQKRLNLQHCYESALDDVPGDVLTEGWARGACITARGILKAY
 DVNRITPVVADSFQGFPRKTIIDDPMDAENNLQYNAVDLPISLAVQNSFRGIGL
 DDQRFPLGWFMDKTPAPRPERLAVLMDSDSYGATWVDLTHYPLRSPGFAITDDY
 CIPCRAGVHRYDRHGIDSEIVIDQGYMRSILGLTFVHCRRSGRSGPTTGACHH
 TNGSSGSPRGNIPSLHRSVCVGSIPSR"
 1490. .2641
 /gene="mycC1"
 1490. .2641
 /gene="mycC1"
 /codon_start=1
 /transl_table=1
 /product="cytochrome P450"
 /protein_id="BACS7023.1"
 /db_xref="GI:28394150"
 /translation="MDRTCAWALPEQYAFRORATGMPAKWDSPTMLVSRHEVRA
 LTVPRVDPPTROPRLSRBDDGDGSRMLMDPSTHRLRMFISAFVQVETMR
 PRKITYADGILDRLLALRPVDLITHALPMSVQICHLGVPEDESRPORESELAS
 RPNDRSNPAILISLBYLDGLVTKTHNPDTGLGTVATERLKGSTTHQELVNNATL
 LLAAGHETSANQTLSTVLTLLRHPETAELRQEPALPNMVDLRLRHSTADGRRA
 TADVLGDHTRADGDITILSSANHDGNTFGAETQIHRPANHVAFGYGPQCIGL
 QNLRLMEVTLIGTLFRVRPALRLAQEPDALRVQGSPIFGIDELVEM"
 2662. .2886
 /gene="mycC1T"
 2662. .2886
 /gene="mycC1T"
 /codon_start=1
 /transl_table=1
 /product="ferredoxin"
 /protein_id="BACS7024.1"
 /db_xref="GI:28394151"
 /translation="MRTVLDERCVGAGQCEATAPBLFTQGDGDLGVDRPVTPLBL
 GPAREAVDRCPVRAIRIESSVRTGMARGAG"
 2883. .3866
 /gene="myd1"
 2883. .3866
 /gene="myd1"
 /codon_start=1
 /transl_table=1
 /product="4-ketoreductase"
 /protein_id="BACS7025.1"
 /db_xref="GI:28394152"
 /translation="MTRDMANRSVVVTGALIGTISHPVEELARAGADVLTGLYSRHR
 AVRELSAHDRLVLLPVDLDERLROVFERBAGAQTIHICALIDENAAVKRASHAG
 ILDAWLRASNLKCNVDGFVEDVLSSAYCGPTASPARBDACRTVRSQNSVGRG
 VLSTITIELIAEOYROYGFGVHLVRRPNVYSPDPSFGSPASRVLPAMIAAEGEGE
 EIKDQGOYTRSFYVTTDLVRALALVETGKRSHLNTTDFEVSMLDLARVVFSTGRT
 ARIHHRKQPVGAPAVLDTTRRSEVDVYTRTLNBELEFVRYRRIRGR"
 3887. .5086
 /gene="mycS"
 3887. .5086
 /gene="mycS"
 /codon_start=1

/transl table=11
/product="methyltransferase"
/protein_id="BAC57026.1"
/db_xref="GI:28394153"
/translation="MTAQTEFEDATYQDVVRLAGHDSLELRLTKYDPMISRLVA
EILSRPSPNDPVLVEALVHGSERFRHFLVVDSPTRPGADGKFGVMTVEYL
TELLRELFGVTRPAGVGTGKLPYLTDBEAEQIGTLLAAGQTEALVAGCSR
KPLSELSSRYFTKFGFLHMTPTHDRHRRDNOQVRLTEIGGVYKHPKMGSSL
BMKSFPPGQIYGLDMOKSHVDLRIITQGDNDARLDRIRARXGPFQVYIDG
SHINAVRISFALLPHRYRPGGIVTEDMTATWPGGQADQESGSLGLKSLI
DAIQHQLPSDPNRSFGYDNRNIVGLHVNNAFVEKGRNDEG1PTW1PRDESLVQ
ASSGGAT"
5181..6443
/gene="mycd"
5181..6443
/gene="mycd"
/codon_start=1
/transl table=11
/product="glycosyltransferase"
/protein_id="BAC57027.1"
/db_xref="GI:28394154"
/translation="MRTALTIGSRGDVQPFILGVGLRARGHDVILGAPESKHLVE
RAGLTVPPTGDDGFFTEEVVDSLRSPYLIDLKALPAVEBXYDQVDEILAAQ
QGADAVVYALITIAAYGRSDVPMMAANPNTWRFPAYESQRLIGPLSLHLF
THARAAREAMRPRVDGRROGLEPGGRSPRLIGREPLYLPSPSVLPKPD
MPARAVTGVFWMDQMOPTRELVDPLADGAPVADVQHLAVHREQLTEAIAA
RAAGRIIVVDGDPVSDVLELHVDYTRMLPRTAAVHGGFGTLEAVNAGPH
VVVVFADHPEFMARLLEIGVAPADVFTLREALVASTVAVKVRBDSIGARARLGE
AVRERGVENACCVSESWIDNSRTLTG"
6720..19643
/gene="mycal"
6720..19643
/gene="mycal"
/note="polypeptide synthase"
/codon_start=1
/transl table=11
/product="protonmagnolide IV synthase 1"
/protein_id="BAC57028.1"
/db_xref="GI:28394155"
/translation="MVPSVTAPEESVALVGSRLPGASGVSEFMDLITTEGRDAVAR
OPDGSWRALAEFTDAGFGMSPROAAADPOQLVLELGEALHEHGVPAITAG
SRGVGVVADYDAYSRLWRTGTPDGYTGLNRSMLNRLSYLELGGSPVVDAS
OSSLYAVHLAGESLRGSDALVAGVSLISTPESTAAAGSLSPGRCPTPAR
ANGVYRGGAAVYVLRSLDAVLDGRVAVHILGAVNNDGSGELADPRAAQEVY
RRAVDSGVPAQVYVTEHGTGTPVGDPEKALCAVAGDADPRADPLHVSXKNI
GHLGAAIGITGLKAVLCVKEVLPPLTLNQTNPGLPLAQLRLVQTVVGRPDAD
GRRILAAVSSFGKGTNAHLVLAQAEBAVRPAVSPTEELPAVPLVLSAREALAGQ
AERLRHLDRHPLRLPDVGFSLATSRDVPKRSVRPAQSPVAAGSGLAEGRVAV
FPGQAGQWGTGMAAASLTESVPAARLTCAAAVTPVMSLIDVILRAVGSPLDRDD
VVOACFAVMSLAELMRSGVPRPAVVGHSQSEIAAAYAGKISLEDARIYALRSQ
LILRELHAGVMSVGLPVAETERLITNMSGRLHLAVNGPNAVAGGDLDELVT
ECERRQARVRAVEVASHPHYERIRDELMTLRVTPQPSVAVSTADVCIITGBE
RLDAGTWNHLQPVRFADTLRQMSDTGTSFVLSAHPMLVLALEILAOETTALC
VYGLSLRDEGQGLRMLASAEAGHGLPDMRKVFGTAGELELPTVAFORVYAP
DAERTVPPAAVDAVYRPAASPEQATPLAATTOQVYDLIRTLVLEAVYQGLPSMET
VDPDLITKQGLSEYAAVERRRLRARKVLSLPATLVYTPPTTAVAPGLGDDLHRE
SQLVITPLVDEPILAVMGCRPFQGSVPEDLMDVAGVDVSPVPRGMPLRE
SLFEDGPDAGTTCYCREGFLDDAAGDFPQGVSPREXAVADPOORLLLEGVREVG
AGFGSAGVGRSTVYAGVMOQYEGRLYESAEHRYGLITGTAASVVGSRVYLG
LGGPATVDTACSSLVALLHVAOLRSECDVALAGVATATPGIPIEFSQGLTA
VDGRCAVSDSADGTGMSRGVYVYVORLSDAVAKRVLAVRGSANVODASNLT
APSGSQQAVITRAALLASRLVGDVUVEGHGTGRLDPIEAOALATTGRRGGSL
LIGSVKSNIGHAQAAAGVAGVAMVAMERVPAVPAHARCAFAQDMSVSAEYVAG
EACAMPETGRPRRAAVSPGVSGTNAHLVLEQAPVPVPAASGAPVSAVADSS
VSPVWLSARSEKGLASAGRLAALSGLSGSGSGSVGVGVDMGSLGSGSVFMR
GVSLSGVRLGVLGAGGVSGLVVGSRVAGSGVVPFGQAGQWGTGMLRGLGEGV
FSRARPVGSGVQGLVDMPRTACREVRGDLGRDVYQISFVPMVLAALRSBSG
VYPRGVHSGQSEITAAVYVSGALSVBGRVYVGRVIGARLGGAGANVWAGSYGV
VBGLLGLGGLGGLAVAVNPGQVAVAGDGKGGVAGAGVAGVAGARLLPVGYSGIV
GLVDVDRLELGLVLEVRGSSVLPYSSVDVGVGVGGLDGGVGNRLRLRFRBEV
VRLVVDGFGSSFVZVSAHPVLVGALEVDVGVGVGVGVGVGVGVGVGVGVGVGV
AAAGVRLGLPQWRVRLVLAGRQVGLPTYRFOHDYVTLMLAPARTVAAAGATTAADA
AGVGMRYVGMKALPVPTPRLHGRVILVGHGATGCVVDLSGVVADILTRHAGTTEHL

Query Match	14.6%; Score 1309; DB 1; Length 58343;
Best Local Similarity	63.6%; Pred. No. 3,7e-85;
Matches 1952; Conservative	0; Mismatches 1107; Indels 8; Gaps 2;
QY 2047	GGGCAAGGGGTCACCTTCTGTCGCGTCCCGCGCGCGCTCGAGAACCAAGTCGCC 2106
DB 48080	GTGCGGCTCGGCTCCACCGGCAACCGGGGTTCGACCGGTGGCGGTACCGCGCAGGA 48139
QY 2107	CTGAGAGATGATCATCCCGGTACGGTTCGACTGGCGGTGACCGCAGCAGCAGCGCTG 2166
DB 48140	GGCGACGAAGGTGTGTGACGGACACACCTCTGTACCCCGGCCCATTCGAGACCGCG 48199
QY 2167	CCTCACTTGACCAATGCTCTCGGACACCGGCACTTCGCGTGTGGCGGCGGTCAACC 2226
DB 48200	CAGGGCCAGTTCACAGGGCGTCAACAGCCTGTTGAC--CGGACGAGATATCCGATCCG 48257
QY 2227	CCGGCGGAATCCGCGCTGAACCCCTTCCGCTCCGACCGCGCGGNNNNNNNNNNNNNN 2286
DB 48258	CAGTAGCGCGCAACTCCGCTCGAAGGCTCGATCTCGGCGCGGACAGTACCAAGCCG 48317
QY 2287	NN 2246
DB 48318	GAGGCGAGCAACCCGCGGACAGGCGCGGTGATCTGCTCGCGCAGCTCTGTGTAAGCCGCT 48377
QY 2347	NCNTCNCCNCCCGCGCGCGGAGGTTCAGCCGCTTCAGACCGCGCGGATTCAGGCC 2406
DB 48378	CCGACGTCAAGAAAGGACCGTCCGCGGTGTGTGGCGGACAAACGCGTCAACGAGGAAGCC 48437
QY 2407	CCGATGCTATCCCAACCCGTCGCGCTCCGCTTCGATTTGATTTGATGCGCGCGGTCAAC 2466
DB 48438	ACGAGCGTCAACCCGTCGCGCGGTGTGTGTAAGTGTGTCAACCGGCGCGTGTGAC 48497
QY 2467	CGACTGATCGAAGACCGTTCAGAGAAAGACTGTCGCCCGCGCTCGATCGATCGTGGCGC 2526
DB 48498	GGCTGTGTGAAACCGTTCATBAATGATCTCTCTCCCGGCTGGGCGCGCATCCGCGCGCC 48557

QY	2527	CGACGTGACGAAGTCGTGGACGACCCGAGTGCAGAGCTCCGGTCCGGGGTGAACCCGGCCGGC	2586
Db	48558	GCTGGCGACGAAGTTCGGGACCACTTCGGCGAGGAGAGTGTCCGGACCGATCCGGCCGGC	48617
QY	2587	GACGTACCGGGTGGCCCCGGCCGAGCCCGGGGAAAACCGAGCTCCCGGTACAGGTAGACGTC	2646
Db	48618	GATTTACCGGGTCGGCACTGGGCGAGTCCGGGAAACCGGAGCTTACGGTTAGAGTTAACGTC	48677
QY	2647	TCCGAGCAGGTGCATCTGCACCCGCACTTGCAGATGGGCGGTTCGGGCGCATGTTTCCGG	2706
Db	48678	GCCGCGAGGTTCACCTGCACCCGCACTTGCAGATGCGGAGATCGCGGACCGATGTCCTCGG	48737
QY	2707	CCGGATCCGCAACGACTGGGCGTGCACCCCGACGACGACTTCCAAACGCTAACCCGAG	2766
Db	48738	CCGGATCCGGATCATCTCCGCGTCCATGCGCCATGACGACATTCAGAGCGCTAACCCGAA	48797
QY	2767	GTCCGCTCGCATGCGCCCGGGTCCGCTTCGGGGGTAGTGCAGCAACCCGGGCGAGCCCTC	2826
Db	48798	GTCCATCTCTCAGCGCTGGGCGCCCGGTGCAGGCGCCGCGCTCGAACTTCCAGAGGAGTC	48857
QY	2827	CCGCAACTCCGGTCCGCTTCGCGCTTCGAGCAACTGCAGCTGCTCCCGACCCGCTTAGTCTC	2886
Db	48858	GCGCAGTTTCGGCGGTTCTGTCGGGGAGAGCGCGCTCCGGTCGGCTCTGTAACTCTC	48917
QY	2887	GCGGACCGGTGACGAAGTTCAGCGCGCTGGCGGACTGGACTCGTTGATTCCGACATGAA	2946
Db	48918	GCGTAACGTTGAGGAAGTTCGATCCGACCGTCCGGGTGCGGCTCGTTAGGTCGGCAGATGA	48977
QY	2947	GTCCACCAAGTTCATGAGCGGTTGGCCCGCCCGCCAGATGATTTGGTTTAGGCCGAG	3006
Db	48978	GTCCACCAAGTTCAGAGCGGCGCGCGCCGACCGGCGAGGACCAAGATGATCATCCCGAG	49037
QY	3007	CCGAGACCGAGTCCGCGCGTTCGGCGGCGCATCCGACGGAAGTCCCGAGAGTTCTTCCGTAC	3066
Db	49038	CCGAGACCGAGCTGCGCGGTTCCGCCCGGAGCTTCTGAAAGTGCAGAGTTGGCGGAC	49097
QY	3067	CCGTTTCGAAAGCGGCAAGCTTGGCCGAGTGGAGGCTGTGTACTGTGCTGTTTCAAGCCGTA	3126
Db	49098	CCGCTTCGAAAGCTTGGAGTTTCCCGTGGTCCGCTGTACTCTCGTGGTCCAGGCGGTA	49157
QY	3127	CAGGGAGGTATGACACCGGCGCCGAGGTCCCAAGGCGGGCTTGGCGGCGCAGAGTCTGTTTC	3186
Db	49158	GAGGCAACCCCGGACCGGCTGCAGAGTCCCAAGGCGCGGTTGATGCGCAGAGCTCTGGTGC	49217
QY	3187	GGTGAGGGCGAAAGCGTTGTGTGTAAGGTCAGGCGGAGGACCCCGCGGGCGCGTGGCA	3246
Db	49218	GGTCAAGCGGAAACGCGTGTGTGTGTAAGTGTACCCGCTGTCTCCCGTGGCGGCGCGGT	49277
QY	3247	CACCAACCGCCCGGATACCCCGGTTGTTCAAACGTTTCCAGGCGCGCGGAGGTACATCCG	3306
Db	49278	CCGCAACGCTCCCGACCTTCGCGGTTGTGTCAAGGCTTCAAGCCCGCCCGGAGAGTACGTGGC	49337
QY	3307	CGTGGGTTTCTCCGACGGGACCTTGTGTATGACCGAGGTGAGGATCTCGTTGCCGGGAC	3366
Db	49338	GTCCGGAATTTGCCCGGCGCACTGTGTGATGACGGAACGAGAGGCGGCGTGCCTGTGTC	49397
QY	3367	CACCGACCTCCCGTGTCTTGGGACCTGTGTACACCGGTACCGAAGGAGTGAACAGCGAAACAT	3426
Db	49398	CAGCAGGAGACGCTGTGTAGACCGCGCCCGCGTACCCCGGACACAGAAAGTGGACGGAAACAT	49457
QY	3427	GCAACTCGGGCGGGGGTGAAGCCCGGACGCTGTACGGGAAAGCGGGCTCCCGCGTACCGG	3486
Db	49458	GCAAGTCAACCCGGGGTGAAGCCCGGACGATCCGGGGGAAACACCGGCTCAAGCCGCAAGCCG	49517
QY	3487	AGCTTCGAACGCGCGGCGGCGGTTCAAGGCGCTGATGTGTGTCTCAAGTACTTTCGCCG	3546
Db	49518	GGCTTCGAACCGGCGACCTGCGTTCAAGCGGAGGACGTTGTGGCGCAAGTACTTTCGCCG	49577
QY	3547	CGGGCCCTCTCAACCGCGCTGGGCCAGCTGGGGATGCATCCAAACAGTTCAGAGCTTC	3606
Db	49578	CGGGCCCTCTCTCAACCGCGCTGGGCCAGCTTCGGGACCTTGGGAAAGATGTTCAGAGCCG	49637

QY	3607	GC	CGAACC	CGG	CCCG	CTG	TCAG	GGT	TCAG	ACC	GAC	CGG	CCG	ATCT	GTCT	CCAA	CGGG	GTGA	CCG	3666	
Db	49638	G	CGAAG	CGCG	CGCG	CGCG	TCAG	CCCG	CGA	GGTTCG	TGCG	GGCG	CTG	CTT	CAGA	GGAGT	GAACG			49697	
QY	3667	CGA	GGCG	CCG	TA	AGT	GGCG	GGT	GA	GTTC	CG	GAC	CGG	AC	CG	CTG	CGTT	CGCG	GGT	TC	3726
Db	49698	T	CGGTT	CC	GA	AGT	TC	AC	GGG	CC	AG	CGG	AC	GG	CG	AG	CGG	CG	CG	CG	49755
QY	3727	GA	TG	CGG	CA	CCG	GT	AG	CGG	GT	GA	CTT	CG	CG	CT	GA	AG	CG	CG	CG	3786
Db	49758	T	T	CG	TC	AG	CG	CC	GA	CG	CG	CG	CT	CA	CG	GG	TA	CG	TA	CG	49817
QY	3787	GT	CGG	CG	TC	GG	TC	GG	CG	TC	CG	GA	TCT	CC	TC	CG	TA	CG	CG	CG	3846
Db	49818	G	T	C	A	G	C	G	T	G	C	G	T	C	G	T	G	A	T	G	49877
QY	3847	G	G	T	C	T	T	A	T	G	-----	A	C	G	T	G	G	A	A	C	3900
Db	49878	C	A	T	T	G	C	C	T	T	T	C	T	G	T	G	A	G	T	C	49937
QY	3901	C	G	T	T	G	C	G	T	T	T	C	C	A	C	T	C	G	G	T	3960
Db	49938	G	C	G	T	G	A	G	CG	GA	CT	G	A	C	T	C	G	A	CG	CG	49997
QY	3961	GA	T	A	AG	T	CG	CA	CA	CC	CG	CG	AG	T	T	C	T	C	A	CG	4020
Db	49998	G	C	A	G	A	T	T	T	C	A	C	CG	CG	CG	CG	CG	CG	CG	CG	50057
QY	4021	CG	AG	CA	CC	CG	TT	CG	CG	CG	CA	C	T	CG	G	T	G	T	CA	CG	4080
Db	50058	C	C	A	T	A	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	50117	
QY	4081	A	G	G	C	G	T	C	A	T	C	G	T	G	C	A	G	G	A	G	4140
Db	50118	A	G	G	C	G	T	C	G	T	G	C	A	G	G	A	G	A	G	A	50177
QY	4141	G	A	G	A	T	CG	CT	CA	CG	CG	GT	C	A	C	G	T	G	C	CG	4200
Db	50178	G	A	G	A	G	A	T	CG	CT	CA	CG	CG	GT	C	A	C	G	T	CG	50237
QY	4201	T	C	A	G	T	A	T	T	G	T	G	T	G	T	G	T	G	T	G	4260
Db	50238	C	C	A	G	T	A	T	T	G	T	G	T	G	T	G	T	G	T	G	50297
QY	4261	C	G	T	C	G	C	A	G	T	C	G	T	G	A	G	C	G	T	G	4320
Db	50298	G	C	C	G	G	C	G	A	G	T	C	G	T	G	A	G	C	G	T	50357
QY	4321	CG	AA	C	CG	T	CG	AG	AG	T	GA	CC	C	A	T	GG	C	C	G	4380	
Db	50358	C	G	A	A	C	CG	T	CG	AG	AG	T	GA	CC	C	A	T	GG	C	50417	
QY	4381	T	G	C	G	A	T	C	G	T	GA	CA	C	T	T	T	G	C	C	CG	4440
Db	50438	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	50477	
QY	4441	G	T	T	C	G	C	A	G	C	CG	T	C	G	T	GA	CA	CC	CG	CG	4500
Db	50478	G	G	G	C	G	C	A	G	C	CG	T	C	G	T	GA	CA	CC	CG	CG	50537
QY	4501	T	G	T	G	C	T	GA	AG	T	GA	CA	C								

Db	50718	AGGGTCGGGGGGTGAATCGCGGCTCCGGACCCGCTTCGGGGTTCGAACAAACCGGTGCGGGGT	50777
QY	4741	CGACGTGCGAGAACACCGGTGTGATGTCGAGCCAGCTCGCCGCTGCGCGTGGCGCGA	4800
Db	50778	CGAGGTGCGAGAACACCGCGCGCGAGCCACGACCTGCGCGGGGTGCGCGTGGCGCGA	50837
QY	4801	AGGTCAATTGACCGGATATCACTTACCCGGTGAAGTCAACGGCCCGAGACCAAGTTTCA	4860
Db	50838	AGGTCAATGAGAGGTACATCACTTCGCGGGTCAATTGCTGGGCCCGCATCAAGATTGCA	50897
QY	4861	GGGCGACGGTGGCGTTCGAGGTGGCGATGAGTGGCGCTAACCCGACCAAGTTCGGACCC	4920
Db	50898	GGGCGACCGTGGCGTTCGAGGTGGCGATGAGTGGCGACCCGACCCGATTCGGCATTC	50957
QY	4921	GGGCGTTCGAACTCCCGACCAAGGGGCCCGCGTGTGGAGACCAAGTTGTTTCAAGGACC	4980
Db	50958	GGTCTCGAATCTCCCGGGTCAAGCGGCCCTCGTGTGGTCAAGCACTGGTGTTCAAAGCCC	51017
QY	4981	ACTTCAGGCGGGCGAGAACCGCTGCGGTCGCGCATCGTCGGCCCGGCCCACTGCAAGG	5040
Db	51018	ATGACACCCGGTTCGAGAACACCGGGCCCGGTCCAAAGGGCGCGGCCGCCCGAAGTGCAGGG	51077
QY	5041	GGTTCAGGAACGCTTCGGGACCCGCCGAAGTTCGCAGATCGTTCGGTAAGCGCTTCAAGC	5100
Db	51078	TCTGGAGAAAGGCTTCCTCCGCGCGCGCAAGAGGGCCAGGTCTGATGTTGCGCTTCAAGG	51137
QY	5101	CGTCGCC	5107
Db	51138	CGCTGCC	51144

3-ketoreductase"
/protein_id="AAFS9931.1"
/db_xref="GI:7329191"
/translation="MPSRLRFFGLGADIALRRTVPALLAHPOVTVAAVSSRTPAR
AAFAAFGEAVPGHQALDRLDDLDALVPLPVWVHPWRAALLRGRHVTVEGDLTA
TRSGADIDIALARSKGLVLMENFTSIHHAQGTITDLDRDTSELSLSAFTTPR
PEGDIRVLPDVGGLLDIGIYPLKAALHPGLDHAAGAVLRERRRRVNSGHWLL
TTPHVVALLAFGEHAYRSYTLFGTARGRLRAAFPEPTTHPRVRIHQDDALLV
DLPDGEANLVIVRDLVAVREGPGRLTQHADAVRQDLVERVNAVARVRC"
complement(1173..2597)
/note="Orf10; involved in the 2-deoxygenation step in
dUDP-L-oleandrose biosynthesis"
/codon_start=1
/transl_table=1
/product="dUDP-4-keto-6-deoxyglucose 2,3-dehydratase"
/protein_id="AAFS9932.1"
/db_xref="GI:7329192"
/translation="MIVGIPLMSRSMGQSVPTAGSEVSTCALLSMIDARRRANRLTVR
HYPFPELQPDQENTQDLRHTSRSPFSEIBELRATDHCMFGSTPQIIVQPELIGLIG
LVKRPDILVHLVQAKNHPGNTIGLQSLPVQATRNATYRVHNGGVRVLYEASPR
GRGRLALVILVQSGSWLTHKRNNMVEALDPLDDDDHWTSLGGTLRTLLRPHVY
NMDRITVYSLCPPEBSPADGROPSAPAAPFAAATVRSITRGKATLHTMGELTGWITDR
SRRELVQQRVPLESTAFSGWRDDHALAHKQDGFYRIGVSVRASRVSWSQPLA
PVGPBLAPVTRIRGVTLVHLIARLTAGLNGEMAPVQCEPLNRAVPAEYRPAVY
LDYVLSADPGRIRYDTLQSEBGRFHHAENRYVVVAEDDPVEVPADPFWLTLHQL
ALHHSNVVVEARSLVACIQALS"
2933..3541
/note="Orf5; involved in the 3,5-epimerization step in
dUDP-L-oleandrose biosynthesis"
/codon_start=1

```

RESULT 14 /c
LOCUS AF237894S1 9523 bp DNA linear BCT 26-MAR-2000
DEFINITION Streptomycetes antibioticus Tu93 oleandomycin biosynthetic gene
cluster, partial sequence.
ACCESSION AF237894
VERSION AF237894.1 GI:7329189
SEGMENT
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

```

protein_id="AAFS9933.1"
/db_xref="GI:7329193"
/translation="MELIDVDGAMLYTPEINRDERGEFLEWFRGRTFOEQEIKGHPPLSLTA
QANCSVSKGVLKRGIIHPADAPPGQAKVTASGVLVDVVDVVRGSGTFCRMAVRID
AARHGGLVLAEGDGHAFMALDADATVYLCSOPVAABERAVPLDPAISIGWPTDID
IVLSKDTHAPSLQAQAAETGILPDYECRCAYIAAAGVBAR"
3538. 4608
/note="OxFe"
/codon_start=1
/transl_table=1
/product="gtdp-D-glucose synthase"
/protein_id="AAFS9934.1"
/db_xref="GI:7329194"
/translation="MKALVLAGSGCTLRITHTSAKQVAVANAKPYLPGLBALAA
GITVGLVGTAGAEVRAAAGVDGAKFGLDITYLEQSPPLAAHAYLIARTYGLDDDEVD
MYLDDGIVAGGIDDLVTRTFRDGRRPARILITVSDPSAFGVELDDDDGQVLEEKY
RHPSPDLAVGVYFFTPALHAAYALIPSRSGELIETHAOLINDADIOSNVITEGY
WKQDGNVADMENVRTVLDLEPRIEGTVEHTVYIGRVVVGGRAGVTRNSINGPAILI
GAGKIGDSYTGPEPTSGDNCRIITGSEMEISIMLABSITGVRIRGSLIGRVQVTO
SLHAPNARFVLGDHSKVEIOS"
4605. 5612
/note="OxFe7"
/codon_start=1
/transl_table=1
/product="gtdp-D-glucose 4,6-dehydratase"
/protein_id="AAFS9935.1"
/db_xref="GI:7329195"
/translation="MNLVLTGAAGFIGSRVYVHLLLEAARGGEPAPVITVLDKLTGAG
VLGAPDDPATVFRGDIADAPLVDLSIMABADQVVPAAESHVRSITSGTFFVRNV
LGTOMLDACLRHGCGFVFNSTDEYVGIIEHSGMPEHOPLCNSPYSAASKASDLA
LSYRTGGLDVARTGRNSNNYGNPHOPKVPALVPTVNLDSLRPLVLDGGIANGEMTILV
DDHCLGVDIVRTGCRGPEGVYHIGGTELTTRDILTGILLDAFGVGMPVDVVPADRKCHD
RRYALDCAKKADELGTRPRRDPABGIAATTIDWTRDNRAWMEPLKKRPAGPAAFRGAAA
P"
5609. 6493
/note="OxFe6; involved in the 4-ketoreduction step in
dtdp-L-oleandrose biosynthesis"
/codon_start=1
/transl_table=1
/product="4-ketoreductase"
/protein_id="AAFS9936.1"

QY 3847 -GGTCTTCATCGACGTGCGAACCCCTTCTGCGCTCTGTGTCGAGGATCAACAACGGTTG 3905
DB 7746 GGGTCTCCCTCCGGAGAGTGGCCCTCTGCGCTCCGCCCTGCTCCGCCGAGGCGCCG 7687
QY 3906 CGTTTCGCTGTTGCC-----ACTCCGCTTGATCAACGCAACCGTGTGTGGC 3955
DB 7686 GGGCTTGACCTGCGCTAGAGCGCGCGCTGGGCGGTGATCCGCGGCCCGTGGCGCGGCG 7627
QY 3956 GAGTCGGATGAGCTGCGACACCGGCGGATGTCCTCATCTGGAACAACGAGGGCGCGTCCG 4015
DB 7626 GACCCGAGATGATTCGACACGCGCGAGATGCTTCGCGGAGAACCGCGGACCGGTGGG 7567
QY 4016 GAGGCGAGACCCGTTTCGCGAGCACTCGCTCTGTGTCAGCCGACGCGTGGTCCGT 4075
DB 7566 CAGGCGAGATGACCTGCTCCGCGAGGTGTTCCGTGTCGCGAGAGAACCGGGGTCTCGGT 7507
QY 4076 GCGGTAGGGGCTCATTTGGTGCGAGCGCGGAGAGAGTGGGCTGTGCGACGACTTCTC 4135
DB 7506 GCGGTAGGGGCTCGCGCTGGTGGCATCCGGGGAGAGAAATAGGGCTGCGCGACGACTTCTC 7447
QY 4136 CGCGCGCAGAGATGCGCTGCGACCGCGTCAACGCTGATGCGGTGGCGGTGCGCTCCACG 4195
DB 7446 CGCGCGCAGAGAGGTGACAGAGAGTCCGCGTGAAGTCCGCTGACCGCGCGCTCATCAG 7387
QY 4196 GATGATCACGTAAGTGTGCTCTCTGTCGCGCGGAGCGAGTGCACGCTGACGCGC 4255
DB 7386 GCGGATGACGTAGTGTGTTGTTGCGTTCGCGCGGTGTGACTGATGACCGGACGCGC 7327
QY 4256 GCGTACGTGCGCAGCTGCTGTGTGAGACCGGTGTGACCCGCTGTGCTCCCTGGT 4315
DB 7326 GGGGAGCGCGAGAGTTCCTGCGCGGTAGAGCTGTAGTTGGCGCGGTGTGCGACCG 7267
QY 4316 CTCGCGGAAACGCTCGAGGAGGTGAGCCCATGCGCGCGGCACTTCGCTCATTTGGC 4375
DB 7266 GTCCGCGAAACGCTCCAGGAGGTGACGCGCCAGTGGCGCGCTCGCTCATTTGGC 7207
QY 4376 GTTGGGCGCGATCTCGGTGACCACTTGTCCGGCGGATGCGGAGTGTGATGATGCGCG 4435
DB 7206 GTTGTATG-----CGGCGCGACCGCGCTGTGCGCGCGGCGGCGGCGCGCG 7153
QY 4436 GATCCGTTGCGCGCAGCGCGCTGTGTGTGACCAACCGCGCGCGCTTCGAGAGCGGTGAC 4495
DB 7152 CAGCGCGTCCGCGCGGTGTGTGTGTGTGACGCGCGCGCGCTTCGAGAGAGTTGAC 7093
QY 4496 CGCTTGT 4555
DB 7092 GACCTTGT 7033
QY 4556 ACCCGTTCGACG 4615
DB 7032 CCGTATGAGT 6973
QY 4616 GCGCATCTTTCGCGCAGCGCTTCACAGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4675
DB 6972 GCGTATGCGCGGAGT 6913
QY 4676 CCGCGCGGTCTGCGCGGT 4735
DB 6912 CCGCGCGGT 6853
QY 4736 CCGGT 4795
DB 6852 CCGGT 6793
QY 4796 CCGGAGGTCTATGTGACGATGATCATCTTTCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 4855
DB 6792 GACGAGAGT 6733
QY 4856 TTTCAGAGGCGAGCGT 4915
DB 6732 TTGAGGTCCCGCGGT 6673

QY 4916 GACCCGCGCGCTGCACTCCCGACACGAGGGGCCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4975
DB 6672 GATGCCGT 6613
QY 4976 GCGCCTTCG 5035
DB 6612 GCGCCTTCG 5035
QY 5036 CAGGCGGT 5095
DB 6552 GAGGCGGT 5095
QY 5096 CAGCGCGT 5155
DB 6492 CATGCTGTCTTTCGCGCGGT 6433
QY 5156 TTGACGT 5215
DB 6432 CCGAGGCGCGAGCG 5215
QY 5216 TCGGGGCGACCGCGCTGT 5247
DB 6372 CCGAGGACGCTGT 6341

RESULT 15
AF055579/C
LOCUS
DEFINITION
AF055579
15052 bp DNA linear BCT 06-JUL-2000
Streptomyces antibioticus putative 3-ketoreductase (oleW),
2,3-dehydratase (oleV), drdp-4-keo-6-deoxyglucose 3,5-epimerase
(oleU), drdp-D-glucose synthase (oleS), dehydratase (oleE),
4-ketoreductase (oleU), dehydratase (oleU), reductase (oleT),
glycosyltransferase (oleT), aminotransferase (oleN2), and
glycosidase (oleT) genes, complete cds.
AF055579
GI:5902165
AF055579.2 GI:5902165
Streptomyces antibioticus
Streptomyces antibioticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 9806 to 15052)
Quiros, L.M., Aguirrezabalaga, I., Olano, C., Mendez, C. and Salas, J.A.
Two glycosyltransferases and a glycosidase are involved in
oleandomycin modification during its biosynthesis by Streptomyces
antibioticus
Mol. Microbiol. 28 (6), 1177-1185 (1998)
9680207
2 (bases 1 to 9805)
Aguirrezabalaga, I., Olano, C., Allende, N., Rodriguez, L., Brana, A.F.,
Mendez, C. and Salas, J.A.
Identification and expression of genes involved in biosynthesis of
L-oleandrose and its intermediate L-olivose in the oleandomycin
producer Streptomyces antibioticus
Antimicrob. Agents Chemother. 44 (5), 1266-1275 (2000)
10770761
3 (bases 9806 to 15052)
Quiros, L.M., Aguirrezabalaga, I., Olano, C., Mendez, C. and Salas, J.A.
Direct Submision
Submitted (25-MAR-1998) Biologia Funcional, Universidad de Oviedo,
Julian Claveria s/n, Oviedo, Asturias 33006, Spain
4 (bases 1 to 15052)
Quiros, L.M., Aguirrezabalaga, I., Olano, C., Mendez, C. and Salas, J.A.
Direct Submision
Submitted (13-SEP-1999) Departamento de Biologia Funcional e
Instituto Universitario de Oncologia de Asturias (I.U.O.P.A.),
Universidad de Oviedo, Oviedo 33006, Spain
Sequence update by submitter
On Sep 16, 1999 this sequence version replaced gi:3023163.
1..15052
Location/Qualifiers
/organism="Streptomyces antibioticus"
/mol_type="genomic DNA"

gene
/acrain="ATCC.11891"
/db_xref="ATCC:11891"
/db_xref="taxon:11890"
complement(590..1576)
/gene="oleu"
complement(590..1576)
/gene="olew"
/function="possibly involved in the C-2 deoxygenation step
in the dtdp-L-oleandrose biosynthesis"
/codon_start=1
/transl_table=1
/product="putative 3-ketoreductase"
/protein_id="A055450.1"
/db_xref="GI:5902166"
/translation="MPSRLRFGVLAGAADIARTPTALAHPOVTVAVASRDTARA
ARPAAGCCAVRGHQAIDLDRIIDALVPLPVVHTVPMVEALIRRHAKLEPLTA
TRSGAEDLIAMARSGVLMENFTSLHAAHGVTYDILIRGTIGELRSLSAATIPPK
PEGDIRYQDPVGGALLDIGIYPLRALHPLGPDLAAGAVALRRERRRVNAGVLL
TPHGVVAELAFGMEHAYRSBYTLFTAGRLRLDRAPETPRHPRVEIRHODALIV
DLPPDAQFANLVRDVLAVRSGPRLTQGHADAVRQADLVERVMAVARVWC"
complement(1596..3020)
/gene="olev"
complement(1596..3020)
/gene="olev"
/function="involved in the C-2 deoxygenation step in the
dtdp-L-oleandrose biosynthesis"
/codon_start=1
/transl_table=1
/product="2,3-dehydratase"
/protein_id="A055451.1"
/db_xref="GI:5902167"
/translation="MIMGIPAMSGVPTPAGSSTCAFLSMLDARRRNLTYE
LVPRRLSGMOFNTGNTLHTSGRFPSTIGLARVTDHCFMGSWTOP1IYQPELIGL
LHYVRFGIHLVYQAKNERNIGLQSLSPVQATRSNYTRVHGSGVRLLEYASR
GRGVLDVLOSBOGSWFLHKGKNNVVEALDDVPLDDFMWISLGGKRLALPHLV
NMDRTVLSCLPDPADPGRPPAPAAFPAAVRSLTRGATALHTGELLGM/LTDR
SRRLVOORVPLESTAFSGWKRDDHAIAHXOGDYFRYIGVSVRASREVSGWQPLA
PVGGLAFTYRIRIGVILVILHARTAGLNGEMAPVQCPRLNRAVADEYRPAY
LDVYLSDPGIRYDTLOSERGGFHHABNRVYVVEADPEVPEVPDPM/LTHQL
ALHHSNYYNRAISVACTOALS"
3356..3970
/gene="olel"
3356..3970
/gene="olel"
/function="epimerization at carbons 3 and 5 in the
biosynthesis of L-oleandrose"
/codon_start=1
/transl_table=1
/product="dtdp-4-keto-6-deoxyglucose 3,5-epimerase"
/protein_id="A055452.1"
/db_xref="GI:5902168"
/translation="WELLIDVNGAMLYTPETMRDRGRTLEWFRGRTQEKIHPISLA
QANCSVRKAPCAASTFTPPGQAKVTCASGVLDVVDVRGSGFTFGMAVRID
AARHQGYLAEGLAHAFMALTDATVYVLCSPYVAEERAVDPLDPAIGLEWPTDID
IVPVGEPTPTRRPWRPRRPGILPDYEGVPGALHRRGGRRGTGP"
3967..5037
/gene="oleS"
3967..5037
/gene="oleS"
/function="activation of glucose-1-phosphate into
dtdp-D-glucose"
/codon_start=1
/transl_table=1
/product="dtdp-D-glucose synthase"
/protein_id="A055453.1"
/db_xref="GI:5902169"
/translation="WKAIVLAGSGSTRLPYTHSAKOLVAVANKPVLYGLEAIAAA
GITVGLIVGTAGVPRVAGDGAQKGLDITRQSRPLGAHVLTAHTVIGDDVY
MYLDNPTVGGITDILVRRPDRPAAITILTHYSDSGVAVLDDGGRVYGEED
RHPKSDLAIVYFTFPAIHEAVRAIEBSMGELEITTAIGHLIDNGADIOSMTYEG
WKDGNVADMLVNRKTVLELEPRIEGVDRHTVTVGVVGSARVANSIMKPAII
GAGPEISDSTYIGPFTSGVGNCRITGSEMEFSLMABSAITVRRIEGSLIGRNVVTO
SLHAPNARFVLGDHSKVEIQS"

gene
5034..6041
/gene="oleB"
5034..6041
/gene="oleB"
/function="4,6-dehydration during 6-deoxysugars
biosynthesis"
/codon_start=1
/transl_table=1
/product="dehydratase"
/protein_id="A055454.1"
/db_xref="GI:5902170"
/translation="MWLVTTGAAGFTGSRVYHLLIENTRGRGPRAPITVLDKLTNG
VLGNVPDPAVTFVRGDIADAPLVDSLMEADVVPFAETHYDRKSTSGTIVKTV
LGTQVLDLALRHGVGPFWNVSTDEVYGSIEHSGWBSHOPICNSPYSAKASDLA
LSYHRTGDLVTRTCNNYGPPOPEKTVPLFVTVMLDLRVLPGDGNLEWMLV
DHCILGVDIRVTRQGRGEVYHIGGTLELTRDITGLLIDFGVMVDVDRKGD
RRYALDCAARAADBLGTRPRDPFABGARTIDWIRDRAMWEP/LKKGPRGAPAAPRGSG
P"
6038..6925
/gene="oleU"
6038..6925
/gene="oleU"
/function="involved in L-oleandrose biosynthesis"
/codon_start=1
/transl_table=1
/product="4-ketoreductase"
/protein_id="A055455.1"
/db_xref="GI:5902171"
/translation="RWLITGAAGMTGREVRLAENEDVVALGHDILVTPSAVR
AALAEHRPGIVAVCAATVADNABETBAAALINAEAPRLAAGLRPHRHGVHLS
DYVPGDARTVADHDPTAPRSAVGRTRKGEAVVATLPAVLAFTAMVIGKTSF
VRVMIERAEAGAIIDVADRGQPTVGTDLADIINAGRPHGILHAATNAGSATVY
DLADEVRLIDAPGRVRRPTTGAFFRAPRPAYSVGHDRMGKTGLAPLDWRMSLR
EAPDILAAHPPTRRGAA"
6922..8103
/gene="oleNT"
6922..8103
/gene="oleNT"
/function="possibly involved in the transamination,
together with olen2, during the biosynthesis of
D-desosamine"
/codon_start=1
/transl_table=1
/product="dehydratase"
/protein_id="A055456.1"
/db_xref="GI:5902172"
/translation="MKRGVYDIALFGDAAFLQPLVYMGPRNTGDRKRLDRLKALNN
RWLNGSLVREFBQRIADLAGVRNVCATNATAGLQILREAVTGEVIMPSMTFYA
TAHAVRWLGVLPVFCIDIDPTGCLDPLVEAAVTPRTGALLGVHLNCRPSRVDLAAI
AAHGLKLFYDAAHALGCTSRORRISFGDAEVSFFATATVVSFEGGGVTDODTBA
ERLRALNFGIGHGVGAGINAKMSAAAGLTSLEAFADAVASNANVELYRQELS
GLQGVRLIDYDPAERNNYVYVILIDAGVGLRDLTLTLRAENVVAOYFSPGCHQ
REPRKTHP/SLHTHEILAOVIALPTGPAVSREDIRKVDIIRVAAHGRPRTAONG
A"
8225..9682
/gene="oleT"
8225..9682
/gene="oleT"
/function="possibly responsible for the 3,4-reduction step
in D-desosamine biosynthesis"
/codon_start=1
/transl_table=1
/product="reductase"
/protein_id="A055457.1"
/db_xref="GI:5902173"
/translation="MALQVTRTPIIDPIAVCAMPGNSAAAAAPLLEBDAPGTGLP
DRLAENHILARRYGTEPFPLEBSARGIGIDRATFVLAIFRTALTRAVRGA
GMWHTNITPLERSGVDAAVRGSPAPVGVGYTPGSCMRCHFCRAVIGARQOQA
LITGONAMFASLIDBMPDNHAYLSGELPLNPGGDIVRRAAGFKLSITYNF
ALTRQGTIDRQGLMDIYALRTSLYGSEDDYVATTTKGAFORKDLTRFOALRRR
RVRPRLGNTITLLEGRGLTGLADYTGDLNDGPPRPVPFTLRREYSGRPDOKAP
BERVELHGAIAFEERIARTAPSLHVDYGYALOSLRGVDALEPRIRPSTRPTAHPO
VAVQVLDLGVYILREAGFPGLQGAERYVAGRLTTGTLEBEVVARFVTEGQVAPRG

Db 7699 CGTCGCGAAGCGCTCCAGAGGAGTCAGGCCCATGCGCGCGCGCTCCGCTCATCTTGG 7640
QY 4375 CGTTGGGCGCATCTCGGTGACCACTTGTCCGGGCGGATCCGAAGTTGTGATGGCCC 4434
Db 7639 CGTTGATG-C-----CGGCCCGACCGCTGTGTGCCCAGGCCGAAGTTGTGACGCGC 7586
QY 4435 GATTCGTTTCGGCCAGCAGGCGCGTGTGTGACCAACCGCCCGCCTCGAAGCGGTGA 4494
Db 7585 GCAGGCGCTCGCGCGCGGT 7526
QY 4495 CCGCTTTGTGGCGGTGAGCTGAACACTTCGCGGTCAACCGATCCGCCACCGACGTC 4554
Db 7525 CGACTTTGTGGCGGTGAGACTGAAGACTTCGCGGTCCGGAAGAGCCGAGCGCGCT 7466
QY 4555 CACCCGTCGTGCAAGCCGAGGCGGTGGCGCGGTGCAAGAAAGCTTGACCTGTGTGTG 4614
Db 7465 GCGGTGAGTGCATCCAGAGGCGGTGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 7406
QY 4615 CCGGATCTTCGCGAGCGCTCCACAGGTGCTGGTCCGCCCAAGGTGTACGCCAGCA 4674
Db 7405 CGGCTATGGCGCGAGTTCTGTCAACCGCGCTGGCGCCGCCAGAGGTGACGCCAGCA 7346
QY 4675 TCGCGCGGTCTTGCGGGGTGAAGAGCGCGGACGTGTCCGGGTCAACAGACCGGTG 4734
Db 7345 TCGCGCGGTGTGGGGGGGTGACCGCGGCTTCAGAGTTTGGGGTCAAGGACGCGGTG 7286
QY 4735 CCGGTTGACGTCCGCAAAACACCGGTGTGATCCAGCAGCTGCGCGCGGTGGCGGTGG 4794
Db 7285 CCGGTTGATGTCCGAAAGACCGGGGCGAGTCCGAGCCACCGGACGCGGTGGCGGTGG 7226
QY 4795 CCGGAAAGTCAATTGACGCGCATGATCACTTCACCGGTGACGTCAACCGGCCGACAGCA 4854
Db 7225 CGAGAAAGTATGACGCGCATGATCACTTCGCGGTGACCTCCGCTTCGCGAGAGCA 7166
QY 4855 GTTCAGGGCGACGCGTGGCGTTGCAAGTGCATGACCGTACCCCGACAGTCCG 4914
Db 7165 GTTGAAGTCCCGGTGGCGTTGCAAGTGCATGACCGTACCGGCGAGTCCG 7106
QY 4915 CGACCCGGGCTCGAATCTCCCGCACAGAGGCGCGCGGTGTGTGAGCAGTTGTGTCA 4974
Db 7105 CGATGCGCTGCTCGAATCTCCCGCACAGAGGCGCGCGGTGTGTGAGCAGTTGTCA 7046
QY 4975 GGGCCCACTCCAGGCGGGCGAGAACCGCTGCGGATCGTCGCGCGCGCCAGT 5034
Db 7045 GGGCCCACTTCAGGCGGTTCAGAGGCGTTTCGCTCACCGGTGTGGGGCGCCCATGT 6986
QY 5035 GCAGGGGTTGAGAAACGCTTCGGGACCGCGAAGATCGCAGATCGGTGCTGACGCT 5094
Db 6985 AGAGGGGCTGAGAGAGCGCGGTGCGCGCGAAGAGGCGCAGGTGCTGTGACGCTGCT 6926
QY 5095 TCACGCGGTGCGCGGT 5154
Db 6925 TCATGCTGCTCTTCGCGCGGTCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6866
QY 5155 GTTGAAGTGTGCTGT 5214
Db 6865 GCGCAGGCGCGAGCGCGCATGACGCAAGGGGTGTCCAGGCGCGTCCGCGCACGCGTGTG 6806
QY 5215 GTCGGGGACCGCGGT 5247
Db 6805 GCCGAGAGCGGT 6773

Search completed: April 7, 2006, 16:58:07
Job time : 43636.7 secs

```

;
OTHER INFORMATION: SEQ ID NO: 6 = tra
NAME/KEY: CDS

```

LOCATION: (5822)...(6595)
OTHER INFORMATION: megDII, daunosaminyl-N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmv hc
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmv homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megBIT-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megAI, KS1
NAME/KEY: misc feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megAI, AT1
NAME/KEY: misc feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megAI, KR1
NAME/KEY: misc feature
LOCATION: (17947)...(18207)
OTHER INFORMATION: megAI, ACP1
NAME/KEY: misc feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megAI, KS2
NAME/KEY: misc feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megAI, AT2
NAME/KEY: misc feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megAI, KR2
NAME/KEY: misc feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megAI, ACP2
NAME/KEY: CDS
LOCATION: (22667)...(33555)
OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (22957)...(24237)
OTHER INFORMATION: megAI, KS3
NAME/KEY: misc feature
LOCATION: (24544)...(25581)
OTHER INFORMATION: megAI, AT3
NAME/KEY: misc feature
LOCATION: (26230)...(26733)
OTHER INFORMATION: megAI, KR3 (inactive)
NAME/KEY: misc feature
LOCATION: (26958)...(27258)

OTHER INFORMATION: megAI, ACP3
NAME/KEY: misc feature
LOCATION: (27353)...(28590)
OTHER INFORMATION: megAI, KS4
NAME/KEY: misc feature
LOCATION: (28897)...(29331)
OTHER INFORMATION: megAI, AT4
NAME/KEY: misc feature
LOCATION: (29953)...(30477)
OTHER INFORMATION: megAI, DH4
NAME/KEY: misc feature
LOCATION: (31396)...(32244)
OTHER INFORMATION: megAI, ER4
NAME/KEY: misc feature
LOCATION: (32257)...(32799)
OTHER INFORMATION: megAI, KR4
NAME/KEY: misc feature
LOCATION: (33052)...(33312)
OTHER INFORMATION: megAI, ACP4
NAME/KEY: CDS
LOCATION: (33666)...(43271)
OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (33780)...(35027)
OTHER INFORMATION: megAI, KS5
NAME/KEY: misc feature
LOCATION: (35385)...(36419)
OTHER INFORMATION: megAI, AT5
NAME/KEY: misc feature
LOCATION: (37068)...(37604)
OTHER INFORMATION: megAI, KR5
NAME/KEY: misc feature
LOCATION: (37860)...(38120)
OTHER INFORMATION: megAI, ACP5
NAME/KEY: misc feature
LOCATION: (38187)...(39470)
OTHER INFORMATION: megAI, KS6
NAME/KEY: misc feature
LOCATION: (39795)...(40811)
OTHER INFORMATION: megAI, AT6
NAME/KEY: misc feature
LOCATION: (41406)...(41936)
OTHER INFORMATION: megAI, KR6
NAME/KEY: misc feature
LOCATION: (42168)...(42425)
OTHER INFORMATION: megAI, ACP6
NAME/KEY: misc feature
LOCATION: (42585)...(43271)
OTHER INFORMATION: megAI, TR
NAME/KEY: CDS
LOCATION: (43268)...(44344)
OTHER INFORMATION: megCI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (44355)...(45623)
OTHER INFORMATION: megCII, desosaminyl transferase, desosamine glycosyltransferase
OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (45620)...(46591)
OTHER INFORMATION: megBIT-2(megBIT), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (46660)...(47403)
OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (47411)...(47980)
OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ
US-09-679-279-1
Query Match 100.0%; Score 600; DB 3; Length 47981;
Best Local Similarity 100.0%; Pred. No. 1.3e-102; Indels 0; Gaps 0;
Matches 600; Conservative 0; Mismatches 0;

QY 1 CCGCGCTCGCGGAGGCTTACACCCGCGGGGTGAGGTGACTGCGCTTACCGCAGTGGGTG 60
Db 36307 CCGCGCTCGCGGAGGCTTACACCCGCGGGGTGAGGTGACTGCGCTTACCGCAGTGGGTG 36366
QY 61 AGGAGAGCGCGCGGTGACCTTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCGG 120
Db 36367 AGGAGAGCGCGCGGTGACCTTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCGG 36426
QY 121 TCCCGCTGGCGCGGTGACCTTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCGG 180
Db 36427 TCCCGCTGGCGCGGTGACCTTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCGG 36486
QY 181 CCGTTCGACCTCGGCGCGGTCTCTCCCTGGCGCGAGCGGCTCTGTGTGTGACCGGAGCGGAG 240
Db 36487 CCGTTCGACCTCGGCGCGGTCTCTCCCTGGCGCGAGCGGCTCTGTGTGTGACCGGAGCGGAG 36546
QY 241 TACCCCGCGCTTGAACGAGCGTGTGCTCGACCGGCTTGAACGAGCGGCGGAGCGGAG 300
Db 36547 TACCCCGCGCTTGAACGAGCGTGTGCTCGACCGGCTTGAACGAGCGGCGGAGCGGAG 36606
QY 301 TGTGTGTGACCGCGGAGTGGCGCGGCGCGGATCGGCGCGGCACTTGCAGCGGCTGAGCGGCA 360
Db 36607 TGTGTGTGACCGCGGAGTGGCGCGGCGCGGATCGGCGCGGCACTTGCAGCGGCTGAGCGGCA 36666
QY 361 CCGCGCTGTGCACTGTGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
Db 36667 CCGCGCTGTGCACTGTGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 36726
QY 421 CCGAGCTGTGACACCTTCTCGCGT 480
Db 36727 CCGAGCTGTGACACCTTCTCGCGT 36786
QY 481 TGT 540
Db 36787 TGT 36846
QY 541 AGGCGAGT 600
Db 36847 AGGCGAGT 36906

RESULT 2
US-07-642-734C-3.
Sequence 3, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/ABed-2 One Abbott
CITY: Park Rd
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652

REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEITICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..1470
OTHER INFORMATION: /function= "approximate span of module 3"
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of module 3"
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of module 3"
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of module 3"
OTHER INFORMATION: beta-ketoacid domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of module 3"
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span of module 4"
OTHER INFORMATION: beta-ketoreductase of module 4"

FEATURE:
NAME/KEY: misc feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatic"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match 47.8%; Score 286.6; DB 2; Length 20235;
Best Local Similarity 69.7%; Pred. No. 1,1e-44;
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

Qy 1 CCGCGCTCGCCGAGCGGTACACCCGGGGGTGAGTGCATCGGCGTACCGGAGTGGGTG 60
Db 13304 CCGCGCTCGCGGAGGCGTACCGGCGGTGAGTGCATCGGCGTACCGGCGTACCGGCGT 13363
Qy 61 AGGAGCGCCCGGTGACCTGCGGCTGACCGGTTCCAGAGACAGAACTTCGAGTCCCGG 120
Db 13364 ACGCGCGCCCGGTGACCTGCGGCTGACCGGTTCCAGAGAGCGGTTACTGAGTCCCA 13423
Qy 121 TCCCTCTGGGCG---CGGATCCCGACACCGGCGAGAGTGGCGTTACCACTCGCTCGGC 177
Db 13424 TCCCGACCGCGCGCGCGCGGACGAGGACGAGAGACGACGCTGGCTTACAGGTGTATGCG 13483
Qy 178 ACCCGCTGACCTCGGCGGCTCTCTCGGCGGAGCGGCTCTGAGTGTGACCGAGCGG 237
Db 13484 GGGAGCGCGAGTGGAGAGCGCTTCTCGGCGGAGCGGCTCTGAGTGTGACCGAGCGG 13543
Qy 238 CAGTACCCCGCGCTGACCGGAGTGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
Db 13544 GCGTGGCGTCCGAGTTGTGAGAGCGCATCCGAGTGTGAGAGAGAGAGAGAGAGAG 13603
Qy 298 TCGTGTGTGACCGCGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
Db 13604 TCTGAGCTCGGAGTGTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13663
Qy 358 GCAACCGCGCTGACCTGAGTGTCTCTGCTCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 13664 CCGAGCGTCTGTGCACTGTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13717
Qy 418 ACCCGAGCTGAGACCTCTGCGTGTGTCCAGGCGGCGGCGGCGGCGGCGGCGGCGG 477
Db 13718 ATCGTGTGTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13777
Qy 478 CCTGTGTGTGTGACCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 537
Db 13778 GCGTGTGTGTGTGACCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13837
Qy 538 CCGAGGCGATGT 597
Db 13838 CGAGGCGATGT 13897
Qy 598 GTG 600
Db 13898 GCG 13900

RESULT 3
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESS: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..7470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: denhydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984

OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12375..13350
OTHER INFORMATION: /function= "approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19145..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

Query Match 47.8%; Score 286.6; DB 3; Length 20235;
Best Local Similarity 69.7%; Pred. No. 1,1e-44;

Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

```
QY 1 CCGGCTGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 60
    |||
Db 13304 CCGGCTGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 13363

QY 61 AGGAGCGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 120
    |||
Db 13364 AGGAGCGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 13423

QY 121 TCCCTGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 177
    |||
Db 13424 TCCCTGCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 13483

QY 178 ACCCGCTGAGCTTGGCGGCTTCTCTCTGCGCGGAGGCTTGGTGTGACCTGCGCTTACCG 237
    |||
Db 13484 ACCCGCTGAGCTTGGCGGCTTCTCTCTGCGCGGAGGCTTGGTGTGACCTGCGCTTACCG 13543

QY 238 CAGTACCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 297
    |||
Db 13544 CAGTACCGCGGAGGCTTACACCCGCGGGGTGAGGTGACCTGCGCTTACCGCACTGGAGT 13603

QY 298 TCGTGTGAGGAGCGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 357
    |||
Db 13604 TCGTGTGAGGAGCGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 13663

QY 358 GCACCGCGCTTGTCCACTGTGTGTCTGTCTGCGCGGAGGCTTGGGCTTGGGCTTGGGCTT 417
    |||
Db 13664 GCACCGCGCTTGTCCACTGTGTGTCTGTCTGCGCGGAGGCTTGGGCTTGGGCTTGGGCTT 13717

QY 418 ACCCGAGCTTGGAGACCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 477
    |||
Db 13718 ACCCGAGCTTGGAGACCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 13777

QY 478 CCTGTGAGCTTGTGACCAAGGAGCGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 537
    |||
Db 13778 CCTGTGAGCTTGTGACCAAGGAGCGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 13837

QY 538 CCGAGGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 597
    |||
Db 13838 CCGAGGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 13897

QY 598 GTG 600
    |||
Db 13898 GTG 13900
```

RESULT 4

```
US-10-042-665A-3
; Sequence 3, Application US/10042665A
; Patent No. 6924106
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Toupet, Christine
; APPLICANT: Engel, Nathalie
; TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
; FILE REFERENCE: 4-21001/B/C1
; CURRENT APPLICATION NUMBER: US/10/042,665A
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: 09/242,744
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: PCT/EP97/04495
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 53799
; TYPE: DNA
; ORGANISM: Amycolatopsis mediterranei
US-10-042-665A-3
```

Query Match 16.3%; Score 98; DB 3; Length 53799;
Best Local Similarity 53.1%; Pred. No. 7.7e-10;

Matches 256; Conservative 0; Mismatches 220; Indels 6; Gaps 2;

```
QY 125 CTTGGGCGCGGAGTCCCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 184
    |||
Db 13599 CTTGGGCGCGGAGTCCCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13658

QY 185 GCACTTGGGAGGAGTCTCTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 244
    |||
Db 13659 GCACTTGGGAGGAGTCTCTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13718

QY 245 CCGGAGCTTGGAGCGAGTGTGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 304
    |||
Db 13719 CCGGAGCTTGGAGCGAGTGTGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 13778

QY 305 GTGACCGGAGGAGTGGCGGCGCGGAGTGGCGGCGCGGAGTGGCGGCGCGGAGTGGCGGCGG 364
    |||
Db 13779 GTGACCGGAGGAGTGGCGGCGCGGAGTGGCGGCGCGGAGTGGCGGCGCGGAGTGGCGGCGG 13838

QY 365 CTTGTCCACTGTGTCTCTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 421
    |||
Db 13839 CTTGTCCACTGTGTCTCTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 13898

QY 422 CAGCTTGGAGACCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 481
    |||
Db 13899 CAGCTTGGAGACCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 13958

QY 482 GTGCTGTGAGACCAAGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 538
    |||
Db 13959 GTGCTGTGAGACCAAGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14018

QY 539 CCGAGGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 598
    |||
Db 14019 CCGAGGCGGAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 14078

QY 599 TG 600
    |||
Db 14079 TG 14080
```

RESULT 5

```
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Park Rd
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckere, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
```

TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc feature
LOCATION: 744..6659
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA"
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc feature
LOCATION: 744..1868
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: /function= "approximate span of
module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: beta-ketoreductase domain/module1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 3831..4811
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6369..6626
OTHER INFORMATION: /function= "approximate span of
module 1"
OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6678..11219
OTHER INFORMATION: /function= "approximate span of
module 2"
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6678..8066
OTHER INFORMATION: /function= "approximate span of
module 2"
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
FEATURE:
NAME/KEY: misc feature
LOCATION: 8262..9305
OTHER INFORMATION: /function= "approximate span of
module 2"
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc feature
LOCATION: 9906..10454

OTHER INFORMATION: /function= "approximate span of
module 2"
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10707..10964
OTHER INFORMATION: /function= "approximate span of
module 2"
OTHER INFORMATION: acyl carrier domain of module 2"
US-07-642-734C-1

Query Match 15.5%; Score 93.2; DB 2; Length 11219;
Best Local Similarity 52.2%; Pred. No. 6.3e-09;
Matches 311; Conservative 0; Mismatches 243; Indels 42; Gaps 3;

QY 1 CGCGCTGCGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 60
DB 9190 CGGCACTGCGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9249
QY 61 AGGAGCGCGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 120
DB 9250 GGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9309
QY 121 TCCCGTGGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 180
DB 9310 CGGAGCGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9369
QY 181 CGGTGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 240
DB 9370 AGGTGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9429
QY 241 TACCGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 300
DB 9397 TGGCGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9456
QY 301 TGTGTGCAACCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 360
DB 9457 AGGTGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9510
QY 361 CGGCGTGTGCAACCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 420
DB 9511 TGGCGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9570
QY 421 CCAGCGTGTGCAACCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 480
DB 9571 GTGCGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9630
QY 481 TGTGTGCAACCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 540
DB 9631 TGTGTGCAACCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTACCGCAGTGTG 9687
QY 541 AGGCGATGTGCGGTGCGGTGCGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTG 596
DB 9688 AGGCGATGTGCGGTGCGGTGCGGCGGAGTACACCGGGGGGTGAGAGTTCATCTGCGGTG 9743

RESULT 6
US-08-439-009A-1
Sequence 1, Application US/08439009A
Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstein
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952, US, D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..6659
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B""
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..1868
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3831..4811
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6369..6626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..11219
OTHER INFORMATION: /function= "approximate span of
```

```
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..8066
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8262..9305
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9906..10454
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10707..10964
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-08-439-009A-1

Query Match 15.5%; Score 93.2; DB 3; Length 11219;
Best Local Similarity 52.2%; Pred. No. 6.3e-09;
Matches 311; Conservative 0; Mismatches 243; Indels 42; Gaps 3;

QY 1 CCGCGCTGCGCGAGGCGTAACCCGCGGCGGTGAGTCACTGCGTACCGCAGTGGGTG 60
DB 9190 CGGACTGCGCGACCGCACCGCGCGGTGCGGTGCGTGGAGAGCGGTGCTCGCGC 9249
QY 61 AGGACCGCGCGGTGACCTGCGCGGTCAACCGGTTCACAGACAGACTTGTGCTCCCGG 120
DB 9250 GGGCGGACGTGTGACCTGCGCGGTATCTTCCAGGAGACGGTTCGCTGCTGCG 9309
QY 121 TCCCCCTGGCGCGGTGCGCGCGCGACCGCGACAGTGGCGTTACAGTGGCGTGGACCC 180
DB 9310 CGGACCGCACCGCTCGTGAACAGTCAACGCGTGTCTTACCGGAGTCACTGACCG 9369
QY 181 CCGTGAACCTGCGCGGTGCTGCGCGCGGACCGGCTCTGTGTGACCGGACCGGACG 240
DB 9370 AGGTGCGCGCTCGGAACCTGCG-----GCGC 9396
QY 241 TACCCCGCGCTGACCGACGTGTGCGCGACCGCGCTGGAACAGCGCGGCGACCGTGC 300
DB 9397 TGCGCGCGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9456
QY 301 TGTGTGTGACCGCGCGAGTGTGCGCGCGCGGATGCGCGCGTCACTGACCGCTCAACGGCA 360
DB 9457 AGGTGCGGTGCG-----CGCTCGCGGACCGCGCGCGGACCGGAGTCAACGCGCGCG 9510
QY 361 CCGCGCTGTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
DB 9511 TCGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9570
QY 421 CCAAGCTGGAACACCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
DB 9571 GTGCGGTGCAAACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9630
QY 481 TGTGCTGTGTGACCAAGGAGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
DB 9631 TGTGACCGGTCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9687
QY 541 AGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596
DB 9688 AGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9743

RESULT 7
US-09-105-537-30
Sequence 30, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
```

```

: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438US1
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: fastseq for Windows Version 3.0
: SEQ ID NO 30
: LENGTH: 13842
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-30.

```

Query Match	14.9%;	Score 89.4;	DB 3;	Length 13842;
Best Local Similarity	52.6%;	Pred. No. 3.1e-08;		
Matches 317; Conservative	0;	Mismatches 271;	Indels 15;	Gaps 5

[illegible]

US-09-105-537-5
 RESULT 8
 ; Sequence 5, Application US/09105537A
 ; Patent No. 6255202
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D. H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.

```

; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600,438U51
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-5

```

Query Match	14.9%	Score 89.4	DB 3	Length 36778
Best Local Similarity	52.6%	Pred. No. 3e-08		
Matches 317, Conservative	0	Mismatches 271	Indels 15	Gaps 5

QY	1	CCGCGCTTCGCGCAGAGCCTTACACCCGCGGGGGGTGAGAGTTCAGCTGCGCTGACCCGACAGTGG---G	57
Db	7590	CCTCGCTTCGCGCAGAGCGGTGGGGTCAAGCGGGGCTTCCCGTGGGATGGACATTTCGCTCTGCGCCG	7645
QY	58	GTAGAGGACGCGCGGTGACCTGTGCGGGTCTAACCCGTTTCACGACGACGACATTCTTGCGCTCC	117
Db	7550	CCAGGCGCTTCGCGCGCCCGGCTTCGCGCCACTTACGCTTTCAGGCCGAGCGGCTACATGCGCTCG	7709
QY	118	CGGTCGCCCTTGCGGCGCGGGATCCCCGACACCGGCGACGAGTGGCGTTTACAGCTCGCGCTGGGC	177
Db	7710	AGAACACTCCCGCC---GCCCTGGCCACCGGCGACGACTGGCGCTTACCGCATTCGACTTGGAA	7766
QY	178	ACCCTGTGACCTTCGCGCGGGTCTCTCCCTGGCGGAGCGGGTCTGTGGTGTGACCGGAGCGCG	237
Db	7767	AGCGCCTTCGCGCGCGCGAGAGGGGTCCGAGCGCACCGG---CTGTCCGCGCGCTGCGCTCG	7822
QY	238	CAGTACCCCGCGCGCTGCGACGAGAGTGTCCGCGGACCGGCGCTTGGAAACAGCGCGGGCGACCG	297
Db	7824	CCGTCAAGCGCGGAGAGACCACTCCGCGGAGCGCGCGCGTGTGTCACCGCGCTGTGTGACG	7883
QY	298	TGCTGTGTGTGACACCGCGCAGATTCGCGCGCCCGAGATCGGCGCGCACTCGACGCGGTGACG	357
Db	7884	CCGGGGGGAAGGTTCGAGGTCTGATCGCGCGGGGCGAGACGACGACCTGTGAGGCCCTTCGCGCG	7943
QY	358	GCAACCGCGCTGTCCACTGTGTCTCTCTGTGCTCGCGCTCGCGAGGAGCGGTGTGACG	417
Db	7944	---CCCGGCTCACCGCACTGACGACCGGTGACGGGCTTTCACCGGCGTGTCTTCGCTCTCG	8000
QY	418	ACCOCAGGCGTGAACCCCTCGCGGTGTGTCCAGAGCGCTTCGCGCGAGCCGCGGATTCGACGTCC	477
Db	8001	ACGGACTGTGACCGGCAAGGTGTCTCGGGTTCGAGGCGCTTCGCGCGACGCCGGAATCAAGCGCGC	8060
QY	478	CCCTGTGGCTGTGACACGAGGAGACGCGCGCGCGCTGACCG---TCGAGACGACGTGCATC	534
Db	8061	CCCTGTGTGTTCGTCACCCGAGGGGGCGGTCTCCGTCCGAGCGTTCGACACCCCGCGCGACAC	8120
QY	535	CGGCGCCAGAGGCTAATGTTCGATGGGCTCGGCGCGGGTGTGGCGTGGAGTCCCCCGCGCGCGT	594
Db	8121	CCGACCGGGGCAATGTCTTGGGGGCTCGGCGCGCGTGTGCGCCCTTGAAGACCCCGGAACGCT	8180
QY	595	GGG	
Db	8181	GGG	8183

RESULT 9
 US-09-120-878-19
 : Sequence 19, Application US/09320878A
 : Patent No. 6117659
 : GENERAL INFORMATION:
 : APPLICANT: ASHLEY, Gary
 : APPLICANT: BERLACH, Melanie C.
 : APPLICANT: BERLACH, Mary C.
 : APPLICANT: MCDANIEL, Robert
 : APPLICANT: TANG, Li
 : TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE


```

; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
; US-09-410-551B-1

```

Query Match 13.3%; Score 79.6; DB 3; Length 77536;

Best Local Similarity 48.6%; Pred. No. 1.9e-06; Matches 248; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

```

QY 1 CCGCGCTCGCGGAGGCGGTACACCGGCGGGGTGAGGTGACGTGCGGTACCGGAGTGGGTG 60
Db 54598 CCGCGCTCGCGGAGGCGGTACACCGGCGGGGTGAGGTGACGTGCGGTACCGGAGTGGGTG 54657
QY 61 AGGAGGC---CCGCTGACCTTGGCGGTCTAACCGCTTCAACGACAGAACTTGTGCTCC 117
Db 54658 GGGACCGCGCGCCGCTGACCGCTCCACGTATCGTTCCAGCAACAAGAACTACTGCTGC 54717
QY 118 CGGTCCCGCGGGCGGGGTCCCGGACACCGGCGGACGAGTGGCGGTACCAAGCTGCGCTGCG 177
Db 54718 GGGCCACCTTCCCGGCGGATGTACCGGCGGCGGCGGACGAGTGGCGGACCGCGCTGC 54777
QY 178 ACCCGCTGACCTCGGCGGTCTCCCTGCGCGGACGAGTCTGTGTGTGACCGGAGCGG 237
Db 54778 TGGGCGCGCGGTGCGCTGCGCGGACCGGCGGAGTCTGTGTGACCGGCGGCTGTGCG 54837
QY 238 CAGTACCCCGCGGCTGAGCGAGCTGTGCTCGGACCGGCTTGAAACAGCGCGGCGGACCG 297
Db 54838 TGGCTTCCATCGGTGCTGCGGAGCACGCGGTGACCGGACCGGTCTCTGCGCGCG 54897
QY 298 TCGTGTGTGACCGGCGAGTCCGCGCGCGGATCGGCGCGGACCTGAGACCGCTGACG 357
Db 54898 CGGCTTCTCTGACCTCGCGCGCGCGCGCGGACGAGTCTGCTGAGCTGTGACG 54957
QY 358 GCACCGCGCGCTGATCAGTGTGTCTCTGTGCTCGCGCTGCGGAGGCGGTGTGACG 417
Db 54958 AACTCTCATTCAGAGCGCGCTGTGTGCTGCGCGGACCGGCGGTGTGCGGTCTCGTGC 55017
QY 418 ACCCGAGCTGACACCTTCGCGGTGTGCTCAGGCGCTGCGGACCGCGGAGTCACTCC 477
Db 55018 AGATCCCGAACCGACGACGCGGCGGCGGCTGACCGTCCACGCGCGGCGGACG 55077
QY 478 CCTGTGTGTGTGACGAGGAGCGCGCGG 507
Db 55078 GCTCGGCGCTGTGAGCCGACGCGCGG 55107

```

Search completed: April 7, 2006, 02:58:55
Job time : 101.965 secs

This Page Blank (uspto)

```
/db_xref="GI:10179841"
```

```
gene
928..2061
/gene="megDVI"
928..2061
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,4-Isomerase"
/db_xref="GI:10179842"
/translacion="WAVGDRRLRGLRGLQARGLYWGANGDLYSMILSGRDDPTTW
YERLRAGRGYASRAGTWWGDHRTAAEVLADGFTHGPPDAARMQVAVHCPAASVA
GPEREYFARTEDASVTVDADWLCQCARLVTELSRFDLVNDFARQVYALGTATA
LKGVDPRLRSVLTATRVCLDAQVSPPOOLATBEQALFALDEIDAVTGDAVAVGVY
AELAAVGNVAVLAVTELPRLAARLADDPETATRVTEVSRTSPGVLERTTASDRR
VGVDVPTGSEVTYVAAANRDEPFTDPDRFDVDRGDAAEILSRPGSRTDLDALV
ATLATAALRAAPVLPRLSRSGVPIRRRSRVAQALSRCPVEL"
2072..3382
/gene="megDII"
2072..3382
/codon_start=1
/transl_table=11
/product="TDP-megosamine glycosyltransferase"
/protein_id="AAG13908.1"
/db_xref="GI:10179843"
/translacion="MRVYFSSMAVNSHLFGVLPPLASAFQAAHGHEVRVVAAPLTDVY
GAGITAPVGDVDELVERMAHAGODIYEMPTLMDWQSHHMSMDLLCMQCTPTPT
FFALMSPDSLIDGVEVFCRMRPMDIYMBPTTAPLPAARVTGTPHARMKGPVATR
ARQSFRLRLAQEVEHREDPLAEWDLTRRFQDPHLSFDBELVLQMTVDPLPEPL
RIDGVTRVGRVYVYNGSVVAPMLLRPERRRVCULTGSSREHGIQGVSIEMLD
AIDIDAEFVAFPDQOLVGVSVAVNRTAGFVPMVILPTCAVTHHGSGTGMELTA
AHHGPOIILSDADTEVAKQLODLGAGLSLPAVGMTAEHLKAGIENYLDSPAVRLGA
ERHEDGKRTDPSPAQVNGICQDLAADAARQRORRTAEPLPR"
3462..4634
/gene="megY"
3462..4634
/gene="megY"
/codon_start=1
/transl_table=11
/product="mycarose O-acyltransferase"
/protein_id="AAG13909.1"
/db_xref="GI:10179844"
/translacion="MVTSTNLDTTARPALNSLTGMRFYAAFLVFTTHVLSRLINSVY
YADIDAMQTTGRCVGSFPIISGFUTGASARSDVSMFRBRVCKLFPNHVTAAR
AAVULPVTGQAVSGBALIIPULLIHAMPALLESFQGNPYSKLACAFYILCPPLF
LFWISGIRBERLMAAVALFAINAVPVADLLPSSPPLIPGLSYAIDQWFLTFP
ATRSLEPFLIGIILARILITGRMINVGLPAVLPFVFNASLFLGVVAISSMMILP
LVLIASGATADLOOKRTFMRNRVVMVLGDVSPALYMHFLVIYVADLLGFSQTEBA
PLGLALFWIIPFLAVSLVLSMLVRFVELPVMRWARPASARRKATPEQTPSRR"
4651..5775
/gene="megDII"
4651..5775
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG13910.1"
/db_xref="GI:10179845"
/translacion="MTTYWSYLLEYERERADILDVOKVAFASGLILGQSENFETE
YARVYGHACVGVNDGTNAVKLALESVGVGRDDEVVTSNTAPVTLAIDEIGARPV
VDVREDYLDMDTDEAVALVPTKAIIVPHVYGCVCMDTALBELADRGLKVECAQ
AHGARDEIRLQAGTSDAASPFPKTVAGVGDGAVVTNDDETARALRLRYGMEE
VYVYTRTGRHSRLDEVGAEILRKULTLDAYVAGRAVAQRVYDGLDSDSHLET
PVVTDGNEHVYVYVVRHPRDEIILKRLIDGDISLNTSYPMVATMGFAHLGASG
SLPVTERRLAGELFSLPMYPSLPHDLDRIEAVREVTIGL"
5822..6595
/gene="megDIII"
5822..6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"

gene
928..2061
/gene="megDVI"
928..2061
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/db_xref="GI:10179846"
/translacion="MRVEELIGISGVPTPTPOTFADRGVGTAYQEDVFAALGRPIF
PVQVSTTRSRGVNRGVHPTTFMGSMARVYCARGAMPADIRGSPFTFGAAEYV
ELASBVMVGLYLPVGMGLFVSLDDTTLVLYLSAGVYPDKERAVHPLDDELALPIPA
DLDLVMSERDRVAPTLREARDQGLIPDYAACRAAHRVAT"
7220..8206
/gene="megDV"
7220..8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translacion="MVTYLGSGFGLGSVNTALADLPYRVRLVAREVYVSGAAYDE
THRDVLTPEALAEVADARAVERPFAAQIDIGTSGMRISBDDVVAERTYNGLVRLIAY
LBSRPAPVAPFSSNTQVGRVAVRQIDSBODHPRVDRQHTQESQDLKEKTAAG
AIRATSLRLPVEFGVPAAGTADRGVSTMI RALQTLQPLTMHDDGVREBELVYTA
ARAFTALDADADLAGRHFLGTRGSRPLGSEFVPAVRSVAREHGEQPVPAVSPPA
HMPDSLRSEVNDPARFTATYCWRAITYTMAEAVDRTYALAPRAAAPSBS"
8228..9220
/gene="megDII"
8228..9220
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AAG13914.1"
/db_xref="GI:10179849"
/translacion="MGTGAGSARVRVGRSALHTSRMLGTVNSGRVTDGDLRLMD
HALERNGICDTADIDYGMRLYKGTBEVGMWFAQGGRRSEVTYLATVYGSBMSRV
DGGISAHNIYAAACNSRLRGVYDIDYOTHTIDRAAPMDVMOAAEHLVSGSGVGY
GSSNLACMHIAAQESARRLNLGHISHQCIYNLAVRPELDVYPAQAIVGVYPAWS
PLHGLLSGVLEKLAAGTAVKASAGRAQVLLPAVRPLVEAYEDYCRRLGADPAEVLGA
WYLSRPGILGAVIQRPTPEQDLSLRAAEVLTGSEBELREBAIFPAPAVDGPVP"
9226..10479
/gene="megBV"
9226..10479
/gene="megBV"
/codon_start=1
/transl_table=11
/product="TDP-mycarose glycosyltransferase"
/protein_id="AAG13915.1"
/db_xref="GI:10179850"
/translacion="MRVLISFARTRTFQGLVPLAMALHTAGHVRVVASQBELTDVYV
GAGLTSVGSDDLFDLISPEAAQVHRYYTDDLPARGRLPRKMERFLHGLEASRPF
VFPPVNNDSFVDREVERFAMDMRPDLVUVEPTFAGAVAAKAGAAHALLMGSDLTGY
FRSSQDLRGORPADDRDPDLGMLTEVAGRFGLDYSBDLAVQMSVQULPESFLFT
GLSEVHRTTLPNGSSVVPOMLRTSGDGRVCFTGGYSALGITSNPOEFLRTLIAR
PDGERIVTTRSGLDASVYDNRALVDVPMNILLPGCAVITHGGAGSGSATLHHGVPO
ISVAHEMDCVLRGKRTAEIAGGVLRPEDEVDADDTLMQALATVEDRSIAENAKLROE
ALAAPTEAVVPEVLEALAHQRADR"
10483..11424
/complement(10483..11424)
/gene="megBIV"
10483..11424
/complement(10483..11424)
/gene="megBIV"
/codon_start=1
/transl_table=11
```

Query Match		100.0%; Score 600; DB 1; Length 47981;	
Best Local Similarity		100.0%; Pred. No. 3e-47;	
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	CCGCGCTCCGCGGAGGAGGTAACCCGGGGGGTGGAGGTCGACCTGCGGACCGGACGAGGCTG 60	/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
Db	36307	CCGCGCTCCGCGGAGGAGGTAACCCGGGGGGTGGAGGTCGACCTGCGGACCGGACGAGGCTG 36366	/protein_id="Agi13916.1"
QY	61	AGGAGCGCCCGGCTGCACTGCGCGGCTACCCGTTCCACAGACAGAACTTCTGGCTCCCG 120	/db_xref="GI:10179851"
Db	36367	AGGAGCGCCCGGCTGCACTGCGCGGCTACCCGTTCCACAGACAGAACTTCTGGCTCCCG 36426	/translation="MTRHVTLGVSGVPSALLREFTTHPLRLRAVARGSPROPS AGIEHLRVLDLBERGVAQVADVDVYVHLVAVAGSTSTRSATVBAEYVAKGIRDS LVAAARPPAPVLLPASTTQANPAERVAQHKEARILIQATEDDVGVILR LPAIVGHSQSGQTRGVVYVAMIRALAGPITMHEGSRVRLMLHVEDAVATLAL HNHEALVGDVMTSPADBARPLGEIPEVVAASVAROTGNPAVPVSVPPENABANDFR SDPDSTEFPTTLTGMHPRVPLAEGIDRTVALISTXE"
QY	121	TCCCGCTGGGCGGGTCCCGGACACCGGCGACGAGTGGCGGTTACAGCTTGGCTGGCACC 180	12181..22821
Db	36427	TCCCGCTGGGCGGGTCCCGGACACCGGCGACGAGTGGCGGTTACAGCTTGGCTGGCACC 36486	/gene="megAT"
QY	181	CGGTGCACTCGGGGGGGTCTCTCCGCGGCGGAGGGGCTCTGGTGGTGAACCGAGCGGCG 240	12181..22821
Db	36487	CGGTGCACTCGGGGGGGTCTCTCCGCGGCGGAGGGGCTCTGGTGGTGAACCGAGCGGCG 36546	/gene="megAI"
QY	241	TACCCCGGCTTGGAGCGAGTGGTCCGCGACCGCTTGGAGACAGCGCGGGGCGACCGTCC 300	/note="polyketide synthase"
Db	36547	TACCCCGGCTTGGAGCGAGTGGTCCGCGACCGCTTGGAGACAGCGCGGGGCGACCGTCC 36606	/codon_start=1
QY	301	TGTTGTGCAACCGGCGAGTGGCGGCGCGGATCGGCGCGGACCTGACGCGCTGACGGCA 360	
Db	36607	TGTTGTGCAACCGGCGAGTGGCGGCGCGGATCGGCGCGGACCTGACGCGCTGACGGCA 36666	
QY	361	CGGCGCTGTCACATGGTGTCTCTCTGCTCGCGGCTCCCGAGGGCGGTGCTTGAAGACC 420	
Db	36667	CGGCGCTGTCACATGGTGTCTCTCTGCTCGCGGCTCCCGAGGGCGGTGCTTGAAGACC 36726	
QY	421	CCAGCTTGGACACCTCGCGGTTGGTCCAGGCGCTCGGCGGACCGGGGATCGACGTCCCC 480	
Db	36727	CCAGCTTGGACACCTCGCGGTTGGTCCAGGCGCTCGGCGGACCGGGGATCGACGTCCCC 36786	
QY	481	TGTGGCTGTGACCAAGGAGCGCGCGCGGTGACCGTCCGAGAGACGAGTGCATCCGAGCC 540	
Db	36787	TGTGGCTGTGACCAAGGAGCGCGCGCGGTGACCGTCCGAGAGACGAGTGCATCCGAGCC 36846	
QY	541	AGGCCATGTGCTCGGTGGGCTTCGCGCGGGTGGTGGGCTGAGTCCCGCGCGGTGGGGTG 600	
Db	36847	AGGCCATGTGCTCGGTGGGCTTCGCGCGGGTGGTGGGCTGAGTCCCGCGCGGTGGGGTG 36906	
RESULT 2			
AX112026		47981 bp	DNA
LOCUS		AX112026	linear
DEFINITION		Sequence 1 from Patent WO0127284.	PAT 01-MAY-2001
ACCESSION		AX112026	
VERSION		AX112026.1	GI:13938925
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
1			
Micromonospora megalomicea			
Micromonospora megalomicea			
Bacteria; Actinobacteriales; Actinobacteridae; Actinomycetales;			
Micromonosporinae; Micromonosporaceae; Micromonospora.			
AUTHORS			
McDaniel, R. and Volchegursky, Y. I.			
TITLE			
Recombinant megalomicin biosynthetic genes and uses thereof			
JOURNAL			
Patent: WO 0127284-A 1 19-APR-2001;			
Kosan Biosciences, Inc. (US)			
FEATURES			
source			
Location/Qualifiers			
1..47981			
/organism="Micromonospora megalomicea"			
/mol_type="unassigned DNA"			
/db_xref="taxon:47868"			
complement(1..144)			
/note="unnamed protein product; megBI (megT),			
TDP-4-keto-6-deoxyglucose 3,4-dehydratase; SEQ ID NO: 2=			
translated amino acid sequence"			
/codon_start=1			
/transl_table=11			
/protein_id="CAC37805.1"			
/db_xref="GI:13938926"			
/translation="MGDRVNGHATPESTQAIRFLTRHGSPPTATDVHDMLHRAAE			
HRLB"			
928..2061			
/note="unnamed protein product; megDI,			
TDP-4-keto-6-deoxyglucose 3,4-isomerase,			
TDP-4-keto-6-deoxyhexose 3,4-isomerase; SEQ ID NO: 3=			
translated amino acid sequence"			
/codon_start=1			
/transl_table=11			
/protein_id="CAC37806.1"			
/db_xref="GI:13938927"			
/translation="MAVGDRRLRGRELQWARGLYMGFGANGDLYSMLSGRDDPMTW			
YRLRAAGCPYASRAGTGVVGDHRTAVALDPGFTTGPDPARMMQVACPAASMA			
GPFREYARTEDAASTVDADMLOORCARLVTEIGSRFDLPNDPAREVPLAGTAPA			
LKGVDPRLSWTSATRVCLDAOVSPQOLAVTEOLATLDEIDAATGAGDAVAVGV			
AEIANTVGNATLATEBELAALADDEFTATRYTEVSRKSPVHLERRTASDRR			
VGVVPTGEGEVTVAAANRDPEVFTDRDQVDGAKILSSRPSPRDLDALV			
ATLATVLAAPAVPLPRLSRGPVIRRRSRPVARGISRCPEVL"			
2072..3382			
/note="unnamed protein product; megDI, rhodoseaminyl			
transferase (eryCII homolog), TDP-megaseaminyl			
glycosyltransferase; SEQ ID NO: 4= translated amino acid			
sequence"			
/codon_start=1			
/transl_table=11			
/protein_id="CAC37807.1"			
/db_xref="GI:13938928"			
/translation="WRVFPSSMAVNSHLFGLVPLASFAQAGHBRVAVASPLTDDVT			
GAGLAVPVGDDVELVEMHAAAGODIVEMRTLDVNDOSHTTMSMDLLGKQTTPT			
FPALMSPSLIDGAVTEPCGSMRPDIWVEPLTAPAPARVGTGPBARMLGDPVATR			
AROSFLRLAHOVEHREBDPLAEFMDTLRRGDPHLSFDESLVGGTVPPIPEPL			
RIDGVRVYGMGVYVNGPSVPAWLLAEPERRVCTLTGSSRRHGIGQVSTGEMLD			
AIDADAEVATFDDQVLGVGVSVPANRTAGFVPMNVILLPTCAATVHNGIGSSWLTA			
AIGHVPQIILSDADTEVHAKQLODLAGSLSPVACMTAEHLGALERTVLDPEAYRIGA			
ERHGDGMTDPSPAVGVICDPLADRARAGRPRTAEPLHPR"			
3462..4634			
/note="unnamed protein product; megG (megY), mycarosyl			
acyltransferase, mycarose O-acyltransferase; SEQ ID NO: 5=			
translated amino acid sequence"			
/codon_start=1			
/transl_table=11			
/protein_id="CAC37808.1"			
/db_xref="GI:13938929"			
/translation="WVSTNLDTTARLPALNSLTGMRFVAPLPIVPTHTVLSRLIPNSVY			
YADGDAFMQTTGRVGVSPRTLSGFTVLTWSBARADSVSPFRRRVCKLPNNHLYTAF			
AAVAVLFTGQAVSGEALPINLLIHAFPALEISFGINPVSWSLACEAFYLCPLP			
LFWISGIRPERMAAAVFAIIVAVPVDLLSPSPPLDGLYSALQDWFVLTTP			
ATRLSLEFLIGIILARILITGRVINGLPAVLLPVPFVASLFLGVVAISSMMILP			
LVLLTASGATADLOOKTFMRVRVVMVGLGVDFALYVMVFLVTVYGADLIGSSQLEDA			
PLGLALFMIPPLAVLSVLSWLVFVELPVMRNARPASARKKATBEQPSNR"			
4651..5775			
/note="unnamed protein product; megDI, deoxy sugar			
transferase (eryCII, DntI homolog),			
TDP-3-keto-6-deoxyhexose 3-amino transferase; SEQ ID NO:			
6= translated amino acid sequence"			
/codon_start=1			

/transl_table=11
/protein_id="CAC37809.1"
/db_xref="GI:13938930"
/translation="MTTYVMSYLLEVERERADITLAVOKYFASGSLILGOSVENFET
YARHGIAHCVCVDNGNTNARKALIESVGRDDEVTYSTAATPYALIDIEIGARPF
VADVEDYLDMDTDLVEAVTPKTAIVFVHLYGQCVMTALRELDKRGKLVEDCQ
AHGARRDGRLAGTMSDAAFSPYTKLAGYGGGAAVYVNDDEFAARLRRLRYGMB
VYVYTRTPGHSRLDEVOAELIRKTLRLDYAGRAVAKORYVDGLADIQDSHGL
PYVDGHEHYFVVVVRPRDEIIRKLDGYDLSINISYPMFWHTMTGPAHGVASG
SLPVTERRLAGRIFSLPMYPSLPHLDQRYVEAREVITG." 5822.
5822. .6595
/note="unnamed protein product; megDIII,
daunosaminyl-N,N-dimethyltransferase (eryCVI homolog) ; SEQ
ID NO: 7= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37810.1"
/db_xref="GI:13938931"
/translation="MPNSHTSTSDVAPYERADIVADFYHGRKGYARADLV
RKHTPOATLIDVACGSHVELADSFREVVDLSAAMLATARDPGRRLHQM
RPSLDVRFPVYVCMPSGTGYLNDRELDRAVNLACHLAPGTLVEPMWPFETPR
GMVGAIDLVTSGDRISRMSTVTPAGLDRTRASRTTIYTGSPAGIEHTVAVMTL
FARAYEAFQFQALSCSYGSHDLFSFGLFVGAABFGR" 6592.
6592. .7197
/note="unnamed protein product; megDIV,
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID
NO: 8= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37811.1"
/db_xref="GI:13938932"
/translation="MRVBEIGTIBGVFTPTPTPADERGVTGTAOXOYFVAALGRPLF
PYAGVSTRRGVYRGVHTFTMGSAKYYTCARGMADPAVDIRGSPTRGGAPEY
ELSAESWGLYLPVGMGHLFVSLDDVTLVLMGAGYVDPKERVHPLDPELALIPA
DLDVMSERDRVAPTLREARDQGLIPDVAACRAAHRVRY" 7220.
7220. .8206
/note="unnamed protein product; megDV, TDP-hexose
4-ketoreductase (eryBIV, dnmV homolog),
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37812.1"
/db_xref="GI:13938933"
/translation="MVVVGASGFLGSAVTHALADIPRVYRIVARREVVVPSGAVADYE
THRVDLTEPGALAEVADARAIFPPAQIRGTSGMRSEDVVAERTNVGLVRLIAY
LSRSPHAAVVVPPGSGNTQVGRVTVIDGSEBQHPBEVYDRKHGTQLKEXATAG
AIRATSLALPVFPGVPAAGTADRGVATIMIRALITGQPLTMHNDGTRELLVYTTA
ARAATVLDHADAALGRHFLIGTGRSMWLGKGFVQAVSRVARSVARGEDVPVPSPPA
HMDPSDLBSVYEDPARFTAVYGMRAITVMAAVDRVTAALAPRAAABSEBS" 8220.
8220. .9220
/note="unnamed protein product; megBII-1(megDVII),
TDP-4-keto-1,6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37813.1"
/db_xref="GI:13938934"
/translation="MGTTGAGSARVRVGRSALHTSRLMIGTVNFSGRVYTDALRLMD
HALRGVACITPADIVGWRLYKGTBELLVGMFAGGGRRETVTLVGSBMSRVN
DGGTSAHIVAAACMSLRIGVDHDIYOTHHIDBAARPELDVLPAAQYVGYRWS
GSSNLGWHIAAQSARIRNLKMTSHOCLYNLAVRPELDVLPAAQYVGYRWS
PLHGLGSLVLEKLAAGTAIVKSAQGRADVLPAVPLVAEYDVRKRGADPAEGLA
WLSRPGIIGAVTGRTPEDQSLRAAELTLGEBELRELAIFPAPVNDPVP" 9220.
9220. .10479
/note="unnamed protein product; megBV, mycascosyl
transferase, mycascosylglycosyltransferase; SEQ ID NO: 11=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37814.1"
/db_xref="GI:13938935"

/translation="MEVLIISFAHRTFGQLVPLMALHTAGHDVAVASQPELTDVV
GAGLTSVPLGSDHRLFDISPEAAQVHARTTTDDFARRKPELSMFLKIEBTSRFE
VPEVANNDSFVDELVEFAMWPRDLTMEFTEFAGVAAKACCAHARLHEDLGTGY
FRSRGODLGGORPADRDPDLGMLTEVAGRFGLDSEDLVQOMVEDQIPESFRLT
GLSEHTRLPLFNGSSVVPOMLRTSGVRCVGTGYSALGITSNOEFRTATLAL
PGEIYVTSGLDPAISPVNRYLVDPVNNILIPGCAVYHGGAGSMATLHGVQ
ISVAHMDKLVKQRTAEIGAGFLEPDEVDATLMQALATVVEDSHENAKLQCE
ALAAPPAEVAPELAEALHQRADR" 12181.
12181. .22821
/note="unnamed protein product; megBIV, TDP-hexose
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;
SEQ ID NO: 12= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37815.1"
/db_xref="GI:13938936"
/translation="MTRVHTLIGSGVPSALLREFTHPLRLRAVARTGSRDOPPS
AGIEHRLVILIEGRVAQVYADTDVYVHLYAVYAGSITMSATVPEAREVNGIMD
LVALARPQPAVLVLPASTQANPAAPSRVIAQHKIEARILROATDEGVDTILR
LPATVYHSGSPGQGTGVVYAMIRALAPGRTIMHSGSVRRMLLVHEDVATFTAL
HNHEALVGVWTSADAEAPLGEIFETVAVSVARQGNPAPVAVSPPPNABANDFR
SDPDSREPTTLGMHPRVPLAGIDRTVALISTXE" 12181.
12181. .22821
/note="unnamed protein product; megA1; SEQ ID NO: 13=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37816.1"
/db_xref="GI:13938937"
/translation="WVDVPLLTGTRTPHPQPLFPFNPPLCGHNPBLRABAROLHAYLE
GISBDDVAVGALALARETRAQDQSPHRAVVASSVTELTALALAQGRPIVGRVLA
RPTAPVFLVPGQAQMPGATRLILAESPVFAACERAPFEBVDMLSTLEVDSBE
HLRREVEVQALFAVQTSIALALRSFEVRDPAVIGSILGILAAVFCGADEVAANA
AALMSREMYPLVAGDMAVALAPAEIALAVERMDVDVAGNPSVLLTGAPEFI
ARVAYELAAGVAAQVAVYVPRFBATRAVILEPGRFTIESPPHVPYLAISIQTLDEVG
RLITRELGAGHFRSFRLEVPFBATRAVILEPGRFTIESPPHVPYLAISIQTLDEVG
SPAIIVPTLQDQGLRFLVAQAATGVYVDMVATVGVTPGHLPSAVAYETDDEG
PSTEFMAADPHYARALREIIVAEALALAGREVDAAATREGLDSVLAVOLTRIA
TATGRDLHIMLYDHPHATLEALRGPBERGREGTAPTEABDEVAVAVAMC
RLPGVTSPEFMEILABGRDAVGLEPTDGMDLDSLPHDPFRSGTAHQAQGLTIG
ATSPDAFRLSAPRELAVERPOORTILELSEVLERAGIPTSLRTSGVYVGLIQ
EYGRILABGSEGVGTMTTSTTSVAGRAYITGLGPRASVDTACSSSLVAVHAC
QSLRGRSTMALAGVAVMPTPMLVDPSMNSLAPGRKAPSAALDAGMGABGACH
LLIERLSDARRHGHPIVAVITGTAIVNSDNGLSAPNGRAQVRVIRQALAESGLTGH

Query Match	100.0%	Score 600	DB 6	Length 47981
Beet Local Similarity	100.0%	Pred. No. 3e-471		
Matches 600	0	Mismatches 0	Indels 0	Gaps 0
1	CCGCGCTCGCGGAGCGTACACCCGGGAGGTGAGGTGCACTGCGCGTACCGCACTGGGTG 60			
Db	36307	CCGCGCTCGCGGAGCGTACACCCGGGAGGTGAGGTGCACTGCGCGTACCGCACTGGGTG 36306		
61	AGGAGCGCGCGGTGACCTGCGCGGTCTACCGGTTCCAAAGCAAGAACTTCTGGCTCCGG 120			
Db	36367	AGGAGCGCGCGGTGACCTGCGCGGTCTACCGGTTCCAAAGCAAGAACTTCTGGCTCCGG 36426		
121	TCCCCCTGGCGCGGTCCCCGACACCGCGGACGAGTGGCGTTACAGTGGCTGGGACAC 180			
Db	36427	TCCCCCTGGCGCGGTCCCCGACACCGCGGACGAGTGGCGTTACAGTGGCTGGGACAC 36486		
181	CCGTGACCTGGGCGGCTCTCCCTGGCGGACGAGGCTCTGATGATGACCGGAGCGGAG 240			
Db	36487	CCGTGACCTGGGCGGCTCTCCCTGGCGGACGAGGCTCTGATGATGACCGGAGCGGAG 36546		
241	TACCCCGCGCTGGAACGAGCTGTGTCGCGACGAGCTTGAACACGCGCGGCGACCGTGG 300			
Db	36547	TACCCCGCGCTGGAACGAGCTGTGTCGCGACGAGCTTGAACACGCGCGGCGACCGTGG 36606		
301	TGTTGTGACCGCGGACGTCGCGCGCGCGGATCGCGCGCTGCACTGACCGCGCTTCAAGGCA 360			
Db	36607	TGTTGTGACCGCGGACGTCGCGCGCGCGGATCGCGCGCTGCACTGACCGCGCTTCAAGGCA 36666		


```

misc_feature      19. .4470
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of module 3'"
misc_feature      97. .1482
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of beta-ketoacyl ACP synthase
of module 3'"
misc_feature      1693. .2670
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of acyltransferase domain of
module 3'"
misc_feature      3406. .3921
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of beta-ketoreductase domain
of module 3', possibly non-functional'"
misc_feature      4171. .4428
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of acyl carrier domain of
module 3'"
misc_feature      4471. .10722
                  /gene="eryA"
                  /note="putative"
                  /function="approximate span of module 4'"
misc_feature      4471. .5847
                  /gene="eryA"
                  /note="putative"

```

misc_feature

/function="approximate span of beta-ketoacyl ACP synthase domain of module 4"

6054..7026

/gene="eryA"

/note="putative"

/function="approximate span of acyltransferase domain of module 4"

7165..9216

/gene="eryA"

/note="putative"

/function="approximate span of dehydratase and enoylreductase domains of module 4"

9433..9984

/gene="eryA"

/note="putative"

/function="approximate span of beta-ketoreductase domain of module 4"

10225..10483

/gene="eryA"

/note="putative"

/function="approximate span of acyl carrier domain of module 4"

10723..20235

/gene="eryA"

/note="ORF3 encoding modules 5 & 6 for 6-deoxyerythronolide B formation; putative"

/codon_start=1

/transl_table=11

/protein_id="AAA26495.1"

/db_xref="GI:152695"

CDS

misc_feature

/function="approximate span of beta-ketoacyl ACP synthase domain of module 4"

6054..7026

/gene="eryA"

/note="putative"

/function="approximate span of acyltransferase domain of module 4"

7165..9216

/gene="eryA"

/note="putative"

/function="approximate span of dehydratase and enoylreductase domains of module 4"

9433..9984

/gene="eryA"

/note="putative"

/function="approximate span of beta-ketoreductase domain of module 4"

10225..10483

/gene="eryA"

/note="putative"

/function="approximate span of acyl carrier domain of module 4"

10723..20235

/gene="eryA"

/note="ORF3 encoding modules 5 & 6 for 6-deoxyerythronolide B formation; putative"

/codon_start=1

/transl_table=11

/protein_id="AAA26495.1"

/db_xref="GI:152695"

Query Match	47.8%	Score 286.6	DB 1	Length 20235
Best Local Similarity	69.7%	Pred. No. 2.2e-18		
Matches 420	Conservative 0	Mismatches 174	Indels 9	Gaps 2
Qy	1	CCGCGCTCCCGAGGCGTACACCCGAGGAGTGAAGTGCATCTGAGGTACCCGAGTGGGTG	60	
Db	13304	CCGCGCTCCCGAGGCGTGAAGGAGGAGGAGTGAAGTGCATCTGAGGTACCCGAGTGGGTG	13363	
Qy	61	AGGAGCGCCCGGTTCGACTTCCGCGGTTCACCGGTTCACAGACAGAACTTCTGGCTCCCG	120	
Db	13364	ACCAGCGGCGCGGTTCGACTTCCGCGGTTCACCGGTTCACAGAGAGAGTGAAGTGGCTCCCA	13423	
Qy	121	TCCCGCTGGGCG---CGGATCCCGACACCGAGGACAGATGGGCTTACAGACTCGACTGGC	177	
Db	13424	TCCCGACCGGCGGAGCGCGACCGAGACAGAGACAGACTGGGCTTACAGAGTCTGATGGC	13483	
Qy	178	ACCACCGTTCGACTTCGAGCGAGTCTCTCCCTTGGCGGACGAGTCTTGGTGTGACCGAGCCG	237	
Db	13484	GGAGACCGAGTGAAGAGAGCGGCTGCTGGCGAGACGATGCTGCTGTGACCGAGCCG	13543	
Qy	238	CAGTACCCCGGCTTCGACAGACATGTGTCCGCGACCGCTTGGACAGCGCGGAGGAGCCG	297	
Db	13544	GCGTGCCTTCGAGTGTGGAAGCCATCCGAAATGATGACTGAGACAGACGAGTGCACCG	13603	
Qy	298	TGCGTGTGTGACCGGCGAGTGCAGCGCCGAGTCCGACCGCACTGACGCGCTTCGACG	357	
Db	13604	TCTGACCTGCGACATGTGATCCCGTTCCACATCGGACACCGCACTGAGGCGCGCGACA	13663	
Qy	358	GACACGCGCTGTCCACTGTGTGTCTCTGTCTGCGCTCGCCGAGGCGGCTGTCTGACG	417	
Db	13664	CCGACGCTGTGTCCACTGTGTGTGTCTGTCTGTCTGCGCGC-----ACGCGAGGCGCTTCG	13717	
Qy	418	ACCCGACGCTTCGACACCTTCGCTTGTGTTCAGGCGCTCGGCGACGCGGATTCGAGTCC	477	
Db	13718	ATCGTGTGTCGACGCGCTGCTCTGTGTTCAGGCGCTTCGAGCGCGCGGATTCGAGTCC	13777	
Qy	478	CCCTGTGCTGTGACCAAGGACGCGCGCGCTTCGACGCTTCGAGACGACGTTCGATCCG	537	
Db	13778	CGCTGTGAGTGTGCTGACCCCGAAGCGCTGAGGTGCGACCGCGAATGTTCGATCCG	13837	
Qy	538	CCGAGGCAATGTTCGTGTGCTTCGCGCGCGGTGTGGCGTGCAGTCCCGCGCGGTGG	597	
Db	13838	CGAGGCGCAATGTGTGGCGGTCTTCGCGCGGTGTGCGCATTCGACGACCGAGGCGCTGG	13897	
Qy	598	GTG 600		
Db	13898	GCG 13900		
RESULT 4				
AR049368	AR049368	20235 bp	DNA	1linear
LOCUS	AR049368	Sequence 3 from patent US 5824513.		PAT 29-SEP-1999
DEFINITION	AR049368			
ACCESSION	AR049368.1	GI:6005407		
VERSION				
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 20235)			
AUTHORS	Katz, L., Donadio, S. and McAlpine, J. B.			
TITLE	Recombinant DNA method for producing erythromycin analogs			
JOURNAL	Patent: US 5824513-A 3 20-OCT-1998;			
FEATURES	Location/Qualifiers			
SOURCE	1..20235			
ORIGIN	/organism="unknown"			
	/mol_type="unassigned DNA"			
Query Match	47.8%	Score 286.6	DB 6	Length 20235
Best Local Similarity	69.7%	Pred. No. 2.2e-18		
Matches 420	Conservative 0	Mismatches 174	Indels 9	Gaps 2

QY	1	CCGGGCTGCGCGAGCCGTAACAACCCGGGGGGGGTGAAGTTCGACTGGCCGTAACCGCAGTGGG	60
Db	13304	CCGGGCTGCGCGAGCCGTAACCGCGAGGGCGTGAAGTTCGACTGGAGCCCGCTTGCCG	13363
QY	61	AGGAGCGCCGGTGCACCTGCGGGTTCACCGCTTCCAGACAGAACTTCTGGCTCCGG	120
Db	13364	ACGGCGGGCCGGTGCAGAGCTGCCGCTTACCGGTTCCAGCGGACAGCGTATCTGGCTGCCA	13422
QY	121	TCCCCCTGGG---CGGGTCCCCGACACCGCGCACAGTGGCGCTTACCACTGGCTGGC	177
Db	13424	TCCCCACCGGAGGGGCGCGCACGGGACAGAGACGACCTGGCGTTACAGTGTATATGGC	13483
QY	178	ACCCTGTGACCTTGAGGCGGTCTTCCCTGGCGGACGGGTCTTGGTGTGACCGGAGCGG	237
Db	13484	GGGAAGCCGATGAGGAGAGCGCTTTCGTGGCGGAGCGCGTGTCTGTGTACCGGACCGG	13543
QY	238	CAGTACCCCGCGGCTGGAACGACGTGGTCCGCGACCGGCTTGAAACAGCGCGGGCGACG	297
Db	13544	GCGTGCAGTCCGATGTTGTTCGAGCGCATCCGAAGTGAATGAGAGAGCGGTCGACCG	13603
QY	298	TTCGTGTGTGACCGCGCAGTTCGGCGCGCCGATTCGGCGCGCACTGACGCGCTGACG	357
Db	13604	TTCGACCTTGGAGGTGAATCCCGTTTCGACATTCGACACCGCACTGAGAGCGCGACGA	13663
QY	358	GCACCGGCGCTTCCACGTGGTCTCTGCTCGCGCTCGCGGAGGGGCGGTCTGTGAGCG	417
Db	13664	CCGACGCTCTGTCACTGTGTGTCTGCTGTCTGCTCCGCG-----AGGGAGAGCCGTG	13717
QY	418	ACCCGACCTTGAGACACCTCGCGTTGTTCAGGCGCTCGCGGACCGCGAGTCAAGTCC	477
Db	13718	ATCCGTGAGTGAACGCGCTCGCCCTGGTTCAGAGGCCCTCGAGAGGGCGGGGTGAAGCAC	13777
QY	478	CCCTGTGGCTGTGACCAAGGAGCGCCGCGCGCTGACCGTTCGAGACAGCTGTGATCCG	537
Db	13778	CGCTGTGGTGTACCCGCAACCCCGTGAAGGTGGCGGACCGCACTGTGTGATCCG	13837
QY	538	CCGAGGCGATGTGCTGGTGGGCTCGGCGGGGTGGGCGTGAAGTCCCGCGCCGGTGGG	597
Db	13838	CGCAGGCGATGTGGGCGGCTCTCGGCGCGGTGTCCGCAATCGACAGACCGGAGGCGCTGG	13897
QY	598	GTG 600	
Db	13898	GCG 13900	
RESULT 5			
AR095529			
LOCUS	AR095529	20235 bp	DNA linear PAT 08-SEP-2000
DEFINITION	Sequence 3 from patent US 6004787.		
ACCESSION	AR095529		
VERSION	AR095529.1	GI:10023473	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 20235)		
TITLE	Katz, L., Donadio, S. and McAlpine, J. B.		
JOURNAL	Method of directing biosynthesis of specific polypeptides		
FEATURES	Patent: US 6004787-A 3 21-DEC-1999;		
source	location/Qualifiers		
	1..20235		
	/organism="Unknown"		
	/mol_type="unassigned DNA"		
ORIGIN			
Query Match	47.8%;	Score 286.6;	DB 6; Length 20235;
Best Local Similarity	69.7%;	Pred. No. 2,2e-18;	
Matches 420;	Conservative 0;	Mismatches 174;	Indels 9; Gaps 2;
Db	13304	CCGGGCTGCGCGAGCCGTAACAACCCGGGGGGGTGAAGTTCGACTGGCCGTAACCGCAGTGGG	60
	13304	CCGGGCTGCGCGAGCCGTAACCGCGAGGGCGTGAAGTTCGACTGGAGCCCGCTTGCCG	13363

OY		61	GGGGAACGCCGTGTGACACTCGCCGGTCTAACCGTTTCAGAACAGAACTCTTGCTCCCG	120
Dd		13364	ACGGCGCGGCCTGTGAGTGTGCCCCCTTAACCCGTTTCAGCGCGCAACGGATCTGGCTGCCA	13424
OY		121	TCCCCCTTAGC---CGGGTCCCAGACAACCGCGCAGAGTGGCGTTTACAGACTCGCTGGC	177
Dd		13424	TCCCCAACGGCGGGCGGGCGAACGGGACAGAGACAGACTGGCGCTTACCAAGTTCGATTGGC	13448
OY		178	ACCCTGTGACCTTCGGGCGGTCTTCCTCTGCGCGGACGGGTCTTGATGTGTGACCGGACGG	237
Dd		13484	GGGAAAGCCGAGTGGGAAAGACGCTTCGTGGCCGGACCGGTCTGCTGGTGAACCGAACGG	13544
OY		238	CAGAACCCCGGCGCTTGACCGGACCGATGTGCTCCGCAACGCGCTGGAAACAGCGCGGGCGACCG	297
Dd		13544	GCGTGGCGGTCCGAGTTGTTCGAGACGCATCCGAACTGTGACCTGAGCACAGAGCGGTTCGACGG	13604
OY		298	TCTGTGTTGTGACACGGCCAGTCTGGCGCGCCCGGATTCGGGCGCGCATCGACCGCTGACG	357
Dd		13604	TCTGTGACTGTGAGACTGTGAATCCCGTTTCAGACATCGGCAACCGCATGTGAGGCGCGCGACA	13664
OY		358	GCACCGCCCTGTGCACTGTGTCTCTCTGTCTCGCGCTCGCCGAGGCGGTGTGTGACG	417
Dd		13664	CCGACGCTGTGTTCACATGTGGTGTGCTGTGCTGTCCCG-----ACGGCGAGCCGTGTG	13718
OY		418	ACCCCAACCTTGAAACACCTTCGCGCTTGTGTCCAGAGCGCTCTGGCGCAGCCGGGATTCAGCTCC	477
Dd		13718	ATCTGTGCTGTGAGACGCGCTCGCTCTGTGTTCAGAGCGCTCTGGAGCGCGGGGTTCAGAGCAC	13778
OY		478	CCCTGTGGCTGTGTGACACGAGGAGACGCGCGCCGCTGTGACCGTCTGAGAGAAGACGTGCATCCGG	537
Dd		13778	CGCTGTGGGTGTGTGACCCCGCAACCCCGTGTGAGGTGGCGCAGACGCGAACTGTGCATTCGG	13838
OY		538	CCCAAGCCATGTGTGTGGGCTGTGGCCGGGTGTGGCGTGTGAATCCCGCCCGGTGTGG	597
Dd		13838	CGCAGGCGCATGTGTGGCGGTCTGTGGCGCGGTGTGTGCGCATTCGAGACCGCGGGCGTGTGG	13898
OY		598	GTG 600	
Dd		13898	GCG 13900	
<hr/>				
RESULT 6				
AY661566		32299 bp	DNA linear	BCT 05-NOV-2004
LOCUS				
DEFINITION		Saccharopolyspora erythraea polyketide synthase gene cluster,		
ACCESSION		AY661566		
VERSION		AY661566		
KEYWORDS		AY661566.1 GI:55419414		
SOURCE				
ORGANISM		Saccharopolyspora erythraea		
		Saccharopolyspora erythraea		
		Bacteria; Actinobacterii; Actinobacteridae; Actinomycetales;		
		Pseudonocardiaceae; Pseudonocardiales; Saccharopolyspora.		
REFERENCE		1. (bases 1 to 32299)		
AUTHORS		Kodumal,S.U., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and		
		Santi,D.V.		
TITLE		Total synthesis of long DNA sequences: Synthesis of a contiguous		
JOURNAL		32-kb polyketide synthase gene cluster		
REFERENCE		Proc. Natl. Acad. Sci. U.S.A. 101 (44), 15573-15578 (2004)		
AUTHORS		2. (bases 1 to 32299)		
TITLE		Hu,Z. and Reid,R.		
JOURNAL		Direct Submission		
		Submitted (21-JUN-2004) Kosan Biosciences, 3832 Bay Center Place,		
		Hayward, CA 94545, USA		
FEATURES				
source		1..32299		
		/organism="Saccharopolyspora erythraea"		
		/mol_type="genomic DNA"		
gene		/db_xref="taxon:1836"		
		1..10638		
		/gene="eryAI"		
		1..10638		
CDS				

/gene="eryAI"
/note="DBS1"
/codon_start=1
/tranal_table=11
/product="EryAI"
/protein_id="AAV51820.1"
/db_xref="GI:55419415"
/translation="MADLSKDSRTAQGRIVRPMPLSGCNESALRARAROLAHLD
RFPDAGVEGGAALAHDEQADAGPHRAVVAASSTSELDLGLAAVADPHASVVRGA
RSPAVYVFPGGAQAWMAAGMAGELIGSRVPAAMDCARAFBYVDWTMLAIVDSRE
QSRREVYQALPAVQTSIALNRSRGVTDAYVGHSTIGLAAHVCGAAGDAANA
AALMSREMIPLVNGDMAVALSDETEPIARMDDVLAGVNGPSVLTSLPEV
ARVQELASGVRAQVINVMAASQVDDIAGSRALMFAQSGSEVFVTSITGG
AVDTRELVAYMRSRFLPYRFDAIRISALEVGTQFEASPHFVLAAALQOQLDAG
SSAAVPTLQRGQGMKRFLLAAQALFTGVAQVDMTAYDDVNGPSLSEPTAAEER
DEPAGSGVDNNAIPHYLRELLVNGETALAGREADARETEPELIVGLAAOLPA
KVSAATGREVNIALIYHPPTPRALAEALAAQETPAQETRAKTEAAGPBRVAVMA
CRLPGVSTPEEFWELISEGDVAAGLPTDRMDLDSLPHDPTRGSTAHQRGGLP
EATFADPAFGMSPREALAVDPQORMLIELMWEVLERAGIPTSLOASPTGVGLIP
QEVGPRLAGSGEGVGLMTGTTSTVASGRLAATLGLGPAISVDTCSSSLVNHIA
COSLRGESSLAMAGVTVMPQMLVDFSPMNSLADPGCKAFSAAGANGMAEGAG
MLLERISDARNGHPVLAVIRGTAVNSDASNGLSAPNGRAOVRVYQALASGGLP
ADIDVAHGTGTTLDGPTIEKARLPAYGRDREPPLHGSYKSNLGTQAAAGVAT
KVLAMAGTLPTTLASERSKELIDSSGALSLDEPEPMAGRPRAAGVSFGISG
TNAAIIEAPQVVEGERVEAGDVVAFWVLSASBGLRAQAALAAHLRHPQDER
DIASLTGRLALPHRAFAFVDESALRVLDGLATGNADGAAGTSRAQORAFFVP
GQGMWGMADVLLDTPVPFAALRECDALPEHLDPEVLPTELEAAREEDALST
BRVNVQPMWAVWVSLASMMRAHGVRAIVITGSHSGEITAAQVAGALSLDDARVVA
LRSRYVITPKNKMASTIADPAGRYRARGRVITAVNGPSSVVAAGDDELRLVA
SCTTECIARRKLVADVASHSHVETIRDALHAELGEDFHPLPGFVPFPSTVGTWTP
DELDAGYRNRLRTVRPADAVRALASQYRTFLEVSAGHPILTAIEEIIIGSGSADLS
AIIHLRGDGSILADPGBALSRFAAGVAVMESYHLTGARVPLPTYPREREVME
PKPARSTREYDEVSAIRYRIEMRPTGAGEPARLDGTMLVAKVAGTAGDESTAREAL
ESAPARRELVYDARCGRDLAERLSRGVAGVLTSLAVDEAREPBARALASLT
LSLYQAVSAELCCPLMTYTESAVATGEFFERYRRAAGALMGCRVIALENPAWGL
VDVPAGVAELARHLAAVSGSAGEDOLALRADSVYGRWRVRAAPATDEWKTGTV
LVYTGTVGGQOILARMLARRGAPHLIVSRGSPDADAGELVAILGALGATVAAAD
VTDRSEVRELLGGIDVPLSAVTHAAATLDDGVDLTGTGERIRASRAKATVAGNHL
ELTEHDLTAFLPSSPASAFAGLGVAGNAVYLDLQOORSDGIPATAVAMGNH
AGSGMAGPVDARPRRHGVIEMRPETACALONALDRAENCPYIYDPRMPTLLAYNA
ORPTRLFDEIDARRALPOAAEPFRVQALASLPAREKALFEIVRSHAAVTLGASA
ERVADQAFELGVDSLALERNRLCAATGVRLPTTVFPHPVRTLAAHLAEIGS
ATGAEOAPATATVADERTIALVQACRLPGEDVSPERIMELITSGRDSAEVPPDRM
VPDELMASDAGTRARAGNFWAGAGDPDAFFGSPREALAMDPOORALLETTEALE
SAGIPETIRGSDTVGVMGSHOGVATGRPRPEVGDGVLITGNTASVSGRIYVIG
LSEPALIYDTRCSSLVALHTRAQSLIDGCGLAVAGSVPMAGPVTPTERSRGALS
PDGCRKPSDEADFGIGESAPVVLQRLSDAREGRVLSVAGSVANDGASGSGS
APSGVAQOVRNIRAMARAGITGADVAVVEAHGTGTRLDPVEASALITATYKSGSSG
PVLGSVKSNIGRAQAAAGVAGVILKVLIGERGVPVPLCRGERSGLIDMSGSEIELA
DGVEMSPADGVRAGVSAFGVGTNAHVITAEPPREPPOPRMLPATGVVPLV
SARTGALIRAOAGRLADHILAAHPCITAPDVUSWTMAROPHREBARVILADTTAAVHR
LRVADGAVVPGVTTGASDGGSVFVFGQAQWEGMARLELPIVPTRESIAECDYVL
SEVAGFSVSEVLERPDAPSLERVDVQVPLFAVMVSLARLMRACGAPSAVITVLSQ
EIIAAVAGALSLERDKMKNVARSRAVAVAVGRGSLVGRSGRDEVLTAADSWTGR
LEVAANVPDAVVVAGDAQARFELECEVGIRARALPVUYASHVAVEVRELVQ
ALAGITPRRAVFPFSTTGDPLTGTELDAGVYRNLRHPVEFHTAVALLODQYAV
IEVSPHYVLTASVOETLDDASDAVYITLERDGDADRPLTALADHTRVAVDMWA
VLGRAGLVLDPGYFPQGRFVLLDORTTPRDELMDGFTRVDTPEPSSEPRALGRML
VVPEGHEEDMTVEVSALAEAGAREPVTGVLGVDGCVVSLALREDAVQTL
VLVRELDABGIDAPLMTVTFGADVAGSVAPADAKLMLGQVSLLEGPGMTGVLVGL
PHMDEPELRGLTAVLAGSEDOVAVADAVARRLSPAHVATSEYAVPGCTIIVTG
TAGAGVARYMLAGRGAEHLVLSRGDTEGVDLTELRLGARVSHVACSSRE
PVRILNGLTIRQGVVGVVAGLPOVALINDMDEAPDRVAVAKAGGAHNLDELCS
DAELFLPSSGAGVWGSARQAYAAKNAPLDAFARHRRGRLPATSVAKGIMAAQMT
GDEBARSFLARSGVAMPVRPALAALDRVLASGETAVVTVTPAPAEASTYARPRP
LILDIRVITPABERAGEPTEBILDRLOGLPABERTAEIVLVTSTAVTLGHDPKAV
RATTPKEIGFDSILAAVLRMLNAAITGLRIPSTLVFHPNAAVAGGLDELCTEYR
GEAPSALAGLALBALPREVPATEREBELVGLERMLAALRPVAAQAAASGCGAMPBGD
DLGRAGVDELLALGRELDDP"

gene
12080.22783
/gene="eryAI1"
12080.22783
/gene="eryAI1"

/note="DBS2"
/codon_start=1
/tranal_table=11
/product="EryAI1"
/protein_id="AAV51821.1"
/db_xref="GI:55419416"
/translation="MTDSERVAVYLRRAVTLIDLRANQRILELSDPIATVSMACRLPG
GVNTPORLMEILREGETLSGFPTDGMWDLARLHHDPDPDPGTSYVDKGFLLDAAAF
DAFFPGSPREALAMPQOQLLEETVENAGIDPSSIRGRTAGVTLCAVAKGXE
DTPAAEDVEGYSATVPAVASGRIISTWMLDEPSSIVDRACSSLVALLAAVSLKE
GESMAYVGGAAVMTQGVDPDVSROGRALADRSRAFGADGFGSGEGLTVILER
LSBARNGEYTLAVRGSALNDQASNGLSAPSGPARQRIITQALSSCGLEPDDVAV
EAGTGALGDPLEANAALDQTRDRADPVLWGSYKSNIGHTQAAAGVGLKTVL
ALNNGELPATLHRETEPTHAGDPVPLVVSARSTALQPMQGRGRTTAAVASGISTGNH
VIVEAREPHERHETTHAGDPVPLVVSARSTALQPMQGRGRTTAAVASGISTGNH
LATTPRHHERRAAVATSTREARVGEIAACDESNAPMDMTKXVDVLRQAPGRDVRVVP
GSMWGMGAEILSSSPVPAKTIACDESNAPMDMTKXVDVLRQAPGRDVRVVP
VLAVWVSLAEILKRSIGVEPAAVVGSHGELAAHAYGALLTEBMAKLVYGRSLNRS
LSGEGMAAVALGEBAVREKRLPMQORLSVAANVNGPSVVVSGEPALRAFSEDCAAE
GIVRDIIDVDVYAGHSPQIERVREBELLETTDIPARPAVTFHSVRESMDGTELDAR
YWRNLRETVRFADAVTRLAESGDVLEIYSPPVVQVAVEAVEADGAEADAVVGS
LHRDGDLSAFSLASMTAHSVGDYIRMDVALPGAAPALPTYEPORRKYVLOPAPAA
ASDELAIVRSWTPIERESGNLGDMLVYTPILSPWMTLCRAINNGRALLRCEVD
TSASRTMAQVADQGTGFGVTSLSDSACRPVPAVAGLITTVQALGDAGVLA
PVNCTQGAVTRADDDLARPAQTVLHGFQVAGLELPGHGGVVDLPSSVDDALNL
LVVLVLRGAEHDLAVRDLRIGRRVVRASLQSGGSRWTPHGTVLVGSAASPVQDQ
LVNMLDRGAREVTLAGACGDLLAAVEBAGASAVVCAQDAALREALDEPVTALV
HAGTINMFGSISVAAPERFAETIAKTALLAVDEVLDGRVAREVYCSGVAIWGA
GMAAYNAGSYLDALAEHHRARSCSTSVAWTPMALPGAVDDQGYLREGLRLSLADR
AMTWERNVLAVGEVAVADVDPVLSBGRPAKTRTALPMLAGRQQAAREPDSGT
GEPARSTAGLSIPBEOQENLELVANAAVEVLGHSAAENVRAPFSELGLDLSLAML
RKLSLSTGRLSPQELSFDPHTVTLAQHILARLVGQDAQAAVAVGASBSEITAY
GIGRFPFGGSGSPQOLMRVLAEGANLTTGPARMQIGRLHYDPNPQTSYVDKGG
FLDDADPDDGFGITPRBALNDPOORMLETMAVEAREVADPDLRGDGVDFG
MNCQSTYQQLAGAEERVDYQIGKNSYLSGRIAYFPGSGEALTYDTRCSSSLVOT
HLMAQILREBESGLAAGVTVMSDPTTFVDFSTQRLADGRCFASBAADGALS
GVVALVLEPSSRRANQVLAVALRGSVAQDASNLAAPNGSQORVIRQALASB
VPADLVLEPVANHGTELDGPIBAGALITAYQODRDEPLTGSKTNIHTQAAAGS
GVILKVLVAMHGMRELHADESEPHIDMSGATVREPEVMPAGPRPAGASGSG
VSGTNAHVITYERAPABOEAARTGRGPPVLTSGSESVVAAQALAEHLRDELEL
TDAANTLATGRARFVPAATVADPDRAGVCAELDALAEGRSADAVPYTSAPRPVTV
FPQGAQWGMQADDLSSSVFABSSRCBALSPTHDMKLDVVRDGGDPRIERVD
VLPVPLFSINVSIAELMRARGVPAAVVAGVSGEIAAANGGTSVAPPTAELEBFPE
VLEHLDQCGMAGSARDELETVLAWMDGVVAVAANGGTSVAPPTAELEBFPE
AABREMKERLIVAYVASHPEVARIEBRLAEIGTITAVGSVYLSHTVGEVDTSA
MDASWYRNLRRLPTEQAVRGIVGEGDFVFEVSPHYLMLVETFAHRAGAETCV
PTLARESGHEFERKILRLAHVHGVGADLRPAVAGSPAEPLPTPESHQFWRPHRP
ADVASLQGERAEHPLLLAAVDPRHGAVFTGRISTDQEWLAEHVAGTILVGSVL
VDLALAGVRAHEHPLLEAVLQRPVLVLAGGALLRMSGAPDESGRATTIVHALEVA
DLADAOQSHATGTLAQVAAAGPDTQMPEDAVRILPDDHYGLAEOGYEYSPFQ
ALDAAWPKDQSVAAEVSIAADESGYAFHVLVLDVACQTLISGLALGEEGGGLPRAMT
VTLHAGSATSVRYVATPAGADMAALRTDPDAGHLVAVVDSLVVRSNGEKKORPBG
EGELHALDMVRLAPSGTGRVADADSLAVLVSGBDDVPVEGAALAMPBGDPRAA
RHEGLHAAVLAWRMLRQEBELPGVLTATGAATVABSDDEVPVEGAALAMPBGDPRAA
ESDPRTFLTDLEPDEPGMLPAVPNPQOLAGDDVFLVREPLPALSATLPGCTIIVLP
GDGALDEVAEPADVDEQPLRAGVRDVAATGDNFVLDLALQMTYPOKMDGTEAAG
VTVAVGEDVDAVFAFGDRLGLFOGAFVAVTTHBRLLAFVDDGSDDAALVAYT
AHYALHDLAELGRAGQVLTIRAAAGVMAVLAARBAQAVLTAAGPAAKIGTILALIL
DDEHTIASRKTGKREKREKTGGGVDTANSLTGELLSDSADLDAEDGVYEMGXTD
LIRAGDGRGVYAPPDLEBADDRLGELIRVAVGLLGRGLDRPVSAMELISLAAVRLG
HMSRGRAVGLVLPDPVDPDGVLTGCTGTGLRLRLVHSHRGLLAVSRG
ADAGSELEAEIETDLGASAEIIACDPTADDAALSALDGLPRPLTVGVAAGVADGL
VTSIDEPAYVOIRAKDQAMNHELFRANTGLSFVLFSSASVLAIPGAGGVVAAANS
SUNALALARTRGIPAKALCWMGMAQASWMTSGSGDRIARTQVALPTERLALFDSA
LIRGGEVVPPLSIRLSALRBAEPVBLRKQVRAKTLAAGDAALAGNVYDRLAGREB
SDQVAGLAEVPLSIRLSAAVSGSADQLEPERAFADGLPDSIAAEILRNLGTAVGRV
PSTLVFHPPTPLAAEHLRDLFAASPAVYIGDILDELKRLSLAESDGHDDVGORL
ESLIRVNSRRADPSTSAISBDSDELSTMLDQRGGGDL"

gene
22784.32299
/gene="eryAI11"
22784.32299
/gene="eryAI11"
/note="DBS3"

```

/codon_start=1
/transl_table=11
/product="EryAIII"
/db_xref="GI:55419417"
/protein_id="AAU51822.1"

Query Match      47.8%; Score 286.6; DB 1; Length 32299;
Best Local Similarity 69.7%; Pred. No. 1.8e-18;
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 1 CCCGCGCTCGCCGAGCGCTACACCCGCGGGGTGAGGTGCACTGCGCTACCGCAGTGGGTG 60
Db 25365 CCGCGCTCGCCGAGCGCTACACCCGCGGGGTGAGGTGCACTGCGCTACCGCAGTGGGTG 25424
QY 61 AGGAGCGCCGCGGTGCACTGCGCTGCTACCGCTTCAACGACGAACTTGTGGCTCCCG 120
Db 25425 ACCGCGGCGCGGTGCACTGCGCTGCTACCGCTTCAACGACGAACTTGTGGCTCCCG 25484
QY 121 TCCCGCTGGGGC---CGGCTCCCGGACACCGGCGACGAGTGGCGCTTACAGCTGGCTGGC 177
Db 25485 TCCCGCTGGGGC---CGGCTCCCGGACACCGGCGACGAGTGGCGCTTACAGCTGGCTGGC 25544
QY 178 ACCCGCTGCACTCGGGCGGTGCTCCCGTGGCGGAGCGGCTGCTGGTGGTGAACGAGGCGG 237
Db 25545 GGGAGCGCGAGTGGGAGAGCGCTTCCGTGGCGGAGCGGCTGCTGGTGGTGAACGAGGCGG 25604
QY 238 CAGTACCCCCCGGCTGCGAGCGAGTGGTCCGCGACGCGCTGGAACAGCGCGGGCGGACG 297
Db 25605 GCGTGCCTGCGAGTGGTGGCGAGCGCATCCGAGTGGAGCTGAGCGAGCGCGTCCACG 25664
QY 298 TCGTGTGTGACACCGCGAGTGGCGCGCGCGGATCCGCGCGCGCACTTGAACGCGCTGACG 357
Db 25665 TCCTGACCTGCGAGTGGATCCCGTTCGACCACTCCGACCGGACCTGAGAGCGCGCGACA 25724
QY 358 GCAACCGCGCTGCACTGGTGTCTCTCTGCTCCGCTCGCGCGAGGCGCGGTGCTGACAG 417
Db 25725 CCAACGCTGCTCACTGGTGTCTCTCTGCTCCGCG---ACGCGAGAGCGCGTTCG 25778
QY 418 ACCCGAGCTGACACCTTCCGCTGGTTCAGAGCGCTCGCGCGAGCGCGGATGACGCTCC 477
Db 25779 ATCCGCTGCTGACAGCGCTCGCTCCGAGCGCTCCGAGCGCGCGGCTGAGAGCAC 25838
QY 478 CCTGTGGCTGTGACCGAGGAGCGCGCGCGCGCTGACCGCTGAGAGCGAGCTGATCCGG 537
Db 25839 CGCTGTGGGTGCTGACCGCGAGCGCGCTGAGCGCGCGCGAGACTGTGATCCGG 25898
QY 538 CCCAGCGCANTGCTCGGTGGCTCGCGCGGTGGTGGCGCTGAGTCCCGCGCGCGGTGG 597
Db 25899 CGCAGGCGATGTGGCGGTCTCGCGCGGTGGTGGCGCATGAGAGCGCGCGCGGTGG 25958
QY 598 GTG 600
Db 25959 GCG 25961

```

```

RESULT 7
SEERYABS
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

SEERYABS
20444 bp DNA linear BCT 28-FEB-1992
S.erythraea eryA gene for 6-deoxyerythronolide B synthase II & III.
X62569
X62569.1 GI:46977
6-deoxyerythronolide B synthase; acyl carrier protein;
acyltransferase; beta-ketoacyl synthase; dehydratase; enoyl
reductase; ketoreductase; multifunctional subunits; thioesterase.
Saccharopolyspora erythraea
Saccharopolyspora erythraea
Bacteria; Actinobacteridae; Actinomycetales;
Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.
Bevitt,D.J., Cortes,J., Haydock,S.F. and Leadlay,P.F.
1
6-deoxyerythronolide B from S. erythraea: cloning of the structural
gene, sequence analysis and inferred domain structure of the
multifunctional enzyme
Bur. J. Biochem.
2 (bases 1 to 20444)
Bevitt,D.J.
Direct Submision
Submitted (19-SEP-1991) D.J. Bevitt, Biochemistry Dept, Univ of
Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK
For related sequences see X56107 & M63677.
Location/Qualifiers
1..20444
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/strain="NRRL 2338"
/db_xref="taxon:1836"
/clone="SE55, DE10, PBR25"
/clone_1lb="genomic in lambda EMBL3"
37..42
46..20268
/gene="eryA"
46..10749
/gene="eryA"
46..10749
/gene="eryA"
/note="ORF B"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase II"
/protein_id="CAA44448.1"
/db_xref="GOA:003132"
/db_xref="UniProt/Swiss-Prot:O03132"
/transl_table="MTDSKVAEYLRATLDDRAAQRIRLESDPIATVSMACRLPG
GNVTPORLWELRSGEETLSGFPTDRGMDLRLHPDPNPGTSTVDKGFDDAGF
DAFPQVSPREAAANDPQRLLETSSWELVENAGIDPHLRGTATVFLGVAFYGE
DTPAAEDVGEVGTGVAIVASGRIISYTMGLGPGSISVTACSSILVALHVESLRK
GSSNAVVGGAIVMTPGVDFPDRORLAAAGRSAPFAGADGFGFSGVTLVLLER
LSEARNRNGHVLAVVRSGLINDGASNGSLASGAPORBITOALSSCGLRBDVAV
EAGCTGACGDIENALIDTIGRDRDNRPLMGSYKSNIGHTQAAAVYGLKLYL
ALNGBELPITLVEEPTPVNDSSGVALLAGNQPWRKERTRAVSAFGISGTVAH
VVERAPERRHETTAHDRPVLVAVASATLARAQAQIAELERPADLADVGLG
LATTRAHERHRAVVAASTEBAVRGLREIAAGAAATDVAVEGTVDGNVAVFLPFGQ
GSGMGAGLSSSPVFAKIRACDESMAPMDKMSVULROAPCAPGLDRVDVVP
VIFAVVSLAELMRSYGVPRPAVVGHSQSLAAVAGALTLDEDAKLVYGRSLMRS
LSGEGMAVAILGEAAVRERLAPMDRLSVAIVNPRSVYSGERPAALAFSEDCAB
GIRVRDIDVIAHSHFQIBRVBELETTGDTAFPRARVTFSTVESRSMGDTEDAR
IYRNRLETRFADAVTRLAESGYDAFLVESHPPVVQAVERAVEADAEADA VVGS
IHRDGDLSAPLRSMATAVSVDIRMDVLAEPALFPTVPRKRYMLPAPAPAA
ADEDELAIVSPTPIKEPESGNDLMTVVTPLISPEWTEMLCEAIINANGRALRCEVD
TSASREMAOVAVAOAGTGRGVLSLSDSDEACRPGVPAGAGLTLVYALDAGVDA
PWCILTQCAVRPPDDDLARPAQTTHAGPAQVAGLEIPRMKGAVUDLPESVDAALRL
LVAIVLGGGRADHLVARDGRLHGRVVASLPQSGRSWTHGTVLVGAASPVGDQ
LVKWLADRGAEHLVLAGACPGDDLLAAVEASAVCAQDDAAALREALGDEPVLAV
HAGTLTNFSGISSEVAPBEERAEFTIAKTALAVLDEVFGDRAVERGVYCSVAGISGCA
GMAAYAGSAVYLDALAEHHRARGSCSTVAATPMALPGAVDGVYLRBGLSLGWA
NARTWERYAAGVAVADVMPVLSSEFAATPTPALPAELAGRGGAABPDGPT
GEPDRIAGLSTPDEQENLLELVANAIVAEVLGHESAAELINVRARPSBELGDSLNMAAL

```

RKRLASTGRLPASTLVDPHTVATLALQHLRAARLVADADQAARVVGAAESEPILAV
GIGRPFPGIGSPGOLMRVLAABGANLTTGPADGMIDIGRLYHDPDNPGRSYVDK
ETLDADPDGPGCTTPRRLALANDPOORLMLKTRAMEVREAGLIDPDALRDTDPVFG
MNGOSTYQQLAGBERVDGTCGLNSASVTSGRATYTFMGEBRLTTDTCCSSIVGI
HLMAOLARKBECSLIAGVTVMSDPTTFVDFSTORGLADGRCAKASADALSB
GVALLVLEPLSRARANGHOVLAVLRGSAVNDQASNGIAAPNGSQRVIROLAAAG
VPADVVAHAGTGLDPIEAGALITATYGODRDRPLGYSKTIHTGTOAAGAG
GIVKIVYAMRGMPLRSLHDEBSPHIDMSGAVLEEVPMFAGSPRAGASVSG
VSGTNAVITYEABARBOEARTEGRLPVLISGSESAVVAOARALAEHLRDEIGL
TDAWNTLATGRAPDVPAATVGDRAVCALDLPLAGRSABAVALVTSAPRPRVTV
FPQGAQWVGAKARDLLESSVEFASMSRCABLSPHDMLKLDVARDGDGDPHERVD
VLQVPLSIWVSLAEWRAGVTPAAVVGSGEIAAHHVAGALSLEAAKVVALLRQ
VIRELIDGQSVAGSADRDELTVLARMDSGVAAVANGPQTSVAGPTAEMLDEFRA
AABEMKPRRIAVVYASHSPVARIBERLALCTITAVNGSVPLHSTVGEVDTA
MDASVWYRNLARPVLFRQAVGVYGEOPDPFVEKSPHVLMAVEBTAHRAAGVITCY
PTLRBQSGPHEPLRMLRAPHYGVADLRPAVAGRPALPTYPHQRPWRPHRP
ADVALVGRGAEPHLLLAADVPGHGAVFTGRISTDBQPLHVVGGVTLVGSVL
VDLAAAGEDVGLPVLBELVLQRPVLVAGAGALRMSGVAPDESGRRTIDVHAADVA
DLADQNSOHAITGLAQVAGAPDTEQWPEDAVRIPLDHYDGLAEQGYEYSPFO
ALRAAMRKDSVVAEVSIAADEBGVAHPVLLDAVQTLISGALGEEGGGLPEAMNT
VTTLASGATSRVIVATPAGADAMALRVTPDAGHVLATVDSLVVSTGKXKROPEPRG
EGELHADMVLAEPGSTGRVVAADSDDAVLSGSEPRPAVTVRTSEBDDPRAA
RHGVLMAALVRRMLRQBELPGATLVATSGAVVSDDSVPBGAAAMGVIRCAQA
ESPRFVLLDTPADPRGMPLPVPDNPOLARGDVFPRLSEPLASALTLPAGTORLVP
GDGAIDSVAFAPADVPQPLRAGEVRVDVRAVGFVDFVLLALGMYPQKADMTGEAG
VTVAVGPDVDAFAPGDVTLGFOCAPAPVATVDRRLARVPGMSDADAAPVLAATY
AHVALHDLAIGRAGQSVLIHAAAGVGMALVLRPAGARVLTARPDGAACGTTPALIT
DDEHTASRETGFARKFRERTGKGVDTVNLSTGLBELDESADLLAEDGVFVEBKTD
LRADGDFRGRYAPPDLEAGADRLGEILREVGLLGAEBDLRLVPSAMELSAALQ
HMSRGRHVGLKVLTOPAPVDPDGTVLITGGTGTGRLLARLVTEHGVRRHLLVSRG
ADAPGSDELRAEIEDIGASAEIACDTPADRLSALDGLPRPLTGVVHAAGVADGL
VTSIDEPABOVILRAKVDAWNLHETLANTGLSPFLFSSAASVLAGGQGVAAAE
SUNLAALARTRGCPARAKLGMGLMAQNSMTSGCDRIARCVGVALPTEPRLALPDA
LRKGEVVPFPLSIRSHAAARFVSVLRKGNVRAKLRAGQAABAGRVNVDRLAGRES
SDQVAGLAEIVSRASAAVSGSADQPERFAFRKALRAAVALAEHLNRILCTAIGVRL
PSTLVDFHPTLAAVAEHLRDLFAASPAVDIGDRLDELKRLKALSLAEDGHDVQORL
BSLRMRSRRPADPSTSAISEDASDELFSMLDQRFEGGSDL"

misc_feature
139..1416
/gene="erya"
/product="beta-ketacyl synthase"
1471..2763
/gene="erya"
/product="acyltransferase"
4261..4455
/gene="erya"
/product="acylcarrier protein"
4513..5787
/gene="erya"
/product="beta-ketacyl synthase"
5839..7116
/gene="erya"
/product="acyltransferase"
7117..7563
/gene="erya"
/product="dehydratase"
8743..9234
/gene="erya"
/product="enoylreductase"
9418..10164
/gene="erya"
/product="ketoreductase"
10315..10509
/gene="erya"
/product="acylcarrier protein"
10750..20268
/gene="erya"
/note="ORF A"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase III"
/protein_id="CAA44449.1"
/db_xref="GI:46979"
/db_xref="GOA:Q03133"

/db_xref="UniProt/Swiss-Prot:Q03133"
/translation="MSGDNQMTREKRLRYRKYRVTREIDSVTALREVERHAGEPIAV
GMACRPGVDVSGSFMFEBRGLIAEAPRAGVSPEDDALGSAAGPDAF
FGISPRBALANDPOORIMLEISWEALREBQDVLSRGATVGTGTGVDGPRPBE
APRVLGTVGTGTPASSVASGRVAVCIAGLBPMATVDGASGGLTALHMEISLRDSC
GLALAGVTVMSPGAFTFRSGGLAADRCRPSFKADPGLIAGACVTLVORLSA
ARREGRPVLAIVAGSVNDQASNGILAPSGQKQVIRTRALNAGVRADVTVARH
GTETRLDPIEVALALSTYGAENGPDDPLTGVSKNTIHTGTOAAGVAKVTVLAR
HGMPTLHEDBSPQIEMPLGASVVSQARSPAEPRPAAVSGSIGTGTNAVTV
EAPRDEPERPADSGPVLVLSGREGORAAQGLADHLAEPRSLADTGTTLAVT
RANAMHRVAVVGDDBALAGLAVALADRILADTTAGCGNPRRVRVYVQALPAAVMS
GMARDLRESQVADSIROGERLAPRVDLILSGAKPRRVRVYVQALPAAVMS
LAAALHSGEPAPAAVVGSGEYLAAHVAAALTLBEAKVAAVRSLRLRLQGGMA
SFLGTBOAERLCRPAALISIVSGPVDVVAAGSGBDELIAECEBAKRAI
PVYASHSPVSELTRELTTELGLIPSVSDVVAAGSTTQPIDTAMAEVYANLR
BOYRPODATHQLAAGDAFVYSPPHVLTVGLATVGLTGCGIGAHLLARASGAEHLV
GLDFFHDLGEAVAGYEVDMSPAPADARVPELVYVPGQYMLPIPGGABRBDOM
RYOVVREABASIASLAGVLLTVGPVSESDALRSGLBQCATVLTCDVSRSTI
GTLAEADDTALSTGVAAVPHGEAVDPSLDALVQALGAQVAVLVTNRNAV
ADELVDPAQAMVGGLRVGIGQPGMGGLVDLVADAASISBLAVLADPGEBOV
AIBADGKVARLVAPARARTHLEPLAGTVLVTGGGIGAHLLARASGAEHLV
LGRGADAQASLRELTALGTVTLACDVADRARLEAVLAAABAEGRVTSAVH
AAVYSTPILDDITAEFTIADVYKRTYVNLDELCPDLDAVTLPSNAGVMSPGIA
SYAANAFTLGGFAPAAASBGAPVTSIAMGLMAQNMAGDGGRTLSQGIKANDPDA
VBEHLITLDGQTSVSVVMDRRPVELFTAAHRRLPDSIAGAPARQSBEGDALR
QRLAALLCDREKREHLHLIRAEVVALTGEDDUALDRDAFPDLGFSMTAVDLNR
LAATVGRBAATVVFPHETTRILADHLETLVGAABEQAAPALVREVPDADPILAI
VGMACRPGVNHMPGELMERIVGGGATVETPDMGLDALPDPDRBGTYSRIG
AFLDGAADPAAFGIISPRALANDPOORVLTETWELFENAGIDPHSVGSDVQVLA
GAAYVQGGQVAPVEBSEGYLLGNSSAVVSGRAVVLGEGPVTDTICSSSLVAL
HSQGLLRDCCGLAVAGVSVNAGPVEFTFSRQGLAVDGRCAKASABADGGLBE
GVAVQGLRSDPBAEGROVLVAVGASAINQOATGLAAPGVAQOQVIRKMAAA
GIVGADVAVVEAGTGTRLDPEYASALITLGTGSRSSGSPVLGYSKTIHQAQAA
GAGVGIKVYGLARGLVPPMLCRGERSPLIEMSGRSLAEVSPPADGVRBACV
SAGVSGTNAHVITIAEPREPIPEBPVYVLAANSVPLLSARTTIAAORLLE
SAVDSPVLTALASALTGSAHLPRKRLALGHSBRLQGLPAAVSGVAPGATGTA
SAGGVVFPQQAQOMEGMARGLISVFPVEBSIAECVAIVSEVAFASVELRQPPA
PSLERVLDVQVPLFSVWVSLARLMGACGPSPVATGHSQSEIAAAVAGVLSLEDGVR
VVLRAKRLALAGKGMVSLAAGBGRALIAEMEDRISVAAVNSGSSVVGSDPRA
LAEIVARCEBGRAKTLLPVDAVSHSHSVREIRITLADIDGISARAAIPLYETLGL
ERKMGERTYTDNRKQVFRDEAVASVSPGHATFVMSHPRTTAAVQETAAVAVI
GSLHRDTAEHLIAELRAHVGAVADMRVPAFPAVLPNTPFEPORITLAEVSD
QLADSRVDMRPLATTPVDLBEGLVHSGAPBELTAVAVEKAVSVVAASDALAL
ALREVPERVAGVSVHTGAANALALHGLISGAGRALMLTSSRAVLAGSEPRADPQ
AMWGLRVMGLETBERMGGLVLDLAPAPADQDAFVACADGDEHOVALRDAARV
RUVYARILGTRESSEMPRAGTALVGTGGVAGRAHRLAAGVDDLVLSRGDAPRA
AAELAEIVALGPKTTITACDVADRBOLSLBELRQGRPVTVVTHATYAGVSEBRPLH
RIGLESVCAKVTGARLDLBCPDATFVLSGAGVMSANATIAHLOALNDGDTCC
HRRABGRAAITSVAMGAMAGGMAVGLBGLTRGRGLPMAVDRALRALHQAUNDGTC
VSIADVMEAPAVAGFTAPRPLIDELVTVAVGAVPAVVOAAPAREMTSQBLLEPTHS
VVAITLHSSPDVAGDOPFTLEPDSILTVAGLRVQALQOATGLPATLVFHPVTRL
ADHTGOOLDGTPARBASALRDGYRQAGSRVSLDLAGISDREHDSGDFS
LDLVMDADGPRSVTVICAGTAIISGHEFTRLAGALRGIAPRVAVQOQVREBEPPL
SSMAVAVQADAVITRQGDPRFVVAHSGAALMAVALATELDRGHPKRVVTLIDY
PQGHQDMANMLBELTATLFDRETVRNDTRLTALGAYDRLTQWRRETEGLPTLLVS
AGEBPMGPMDSDSMPTWPFSDHTVAVPDHFTVQEHADAIARHIDMLGSGNS"

misc_feature
10867..12111
/gene="erya"

Query Match 44.9%; Score 269.2; DB 1; Length 20444;
Best Local Similarity 68.0%; Pred. No. 8.5e-17;
Matches 408; Conservative 0; Mismatches 183; Indels 9; Gaps 2;

QY 1 CCGGCTGCGCGAAGCGTCAKACCCGGGGGGTGGAGTCTGACTGGCGTACCGGAGTGGGTG 60
Db 13337 CCGGCTGCGGAGGCGTACCGGAGGCGTGGAGGTGACTGGAGCCGCCCTTCGCG 13396
QY 61 AGGACGCCCGATGCACTGCGCGGTCTACCGGTTCACAGCAACAAGACTTGTGCTCCCG 120
Db 13397 ACGCGCGCGGTGCACTGCGCGGTCTACCGGTTCACAGC---GTACTGCGTGGCCATCC 13453
QY 121 TCCCGTGGGCGGGTCCCGACACCGGCGACGAGTGGCGTTACCACTCGCGTGGCACCC 180

Db	13454	COACCGGCGGGCGGACACGGGACACAGACACACTGGCGCTTACAGCTGTATGGCGGG	13513
Qy	181	CCGTGACCTTCGGGCGGCTCTCCTTGGCCCGACCGAGTCTCTGTGTGTGACCGGACCGGACG	240
Db	13514	AAGCCGAGTGGAGAGCGCTTCCTCGTGGCCGGAACGCTGTGTGTGACCGGACCGGCGG	13573
Qy	241	TACCCCGGCGCTGGACCGGACGTTGGTTCGGCGACCGGCTTGGAACAGCCCGGGCGGACCGTGG	300
Db	13574	TGCGGTCCGAGTGTTCGAGCCCACTCCGAAGTGAATCTGAGAGCAGAGCGGTGCGACCGTTC	13633
Qy	301	TGTTGTGACACCGCGCACTGCGGCGCGCCGATTCGGCGCGCGCACTGACCGCGTTCGACGGCA	360
Db	13634	TGACCTGGGAGAGTGGAGATCCCGTTTCGACACATCCGCAACCGGACTGGAAGCCCGGACACCG	13693
Qy	361	CCGCCCTGTCCACTGTGAGTCTCTCTGCTGCGCGCTTCGCGGAGCGGTCTGTTCGACGACCC	420
Db	13694	ACGCTGTGTCCACTGTGGGATG-----TCCGTGTGTGCCGACCGGCGAGGCGGTGTGATTC	13747
Qy	421	CCAGCTTGGAACACCTTGGCGTTGGTTCAGAGCGCTTCGGCGGACGCGCGGATTCGACGTCCCC	480
Db	13748	CGTGGCTGGAAGCGCGCTCGCCCTGTGTTCAGGCGCTTCGAGAGCGCGGGGTTCGAAGCACCCG	13807
Qy	481	TGTGAGTGTGACACGAGGAGACGCGCGCGCGCTGACCGCTCGGAGACGACGTTCATTCGGGCC	540
Db	13808	TGTGGGTCTGATCCCGCAACGCGCTGTGAGGTGGCCGACGCGGACACTGTGTTCATTCGGCGC	13867
Qy	541	AGGCCATGTGTCGTGTGGGCTCGGCCGGTGTGTGGGCGTGAAGTCCCGCCCGGTGGGATG	600
Db	13868	AGGCCATGTGTGGGCGGTCTCGGCCGCGGTGTGTGGCATTCGACGACCGGGGCGCTGGGGCG	13927
RESULT 8			
SEERYA			
LOCUS	9769 bp	DNA	linear BCT 14-OCT-1991
DEFINITION	S. erythraea eryA gene for 6-deoxyerythronolide B synthase.		
ACCESSION	X56107		
VERSION	X56107.1 GI:46975		
KEYWORDS	eryA gene; erythromycin.		
SOURCE	Saccharopolyspora erythraea		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.		
REFERENCE			
AUTHORS	1 Cortes,J., Haydock,S.F., Roberts,G.A., Bevil,D.J. and Leadlay,P.F.		
TITLE	An unusually large multifunctional polypeptide in the		
JOURNAL	erythromycin-producing polyketide synthase of Saccharopolyspora		
PUBMED	erythraea		
REFERENCE	Nature 348 (6297), 176-178 (1990)		
AUTHORS	2 (bases 1 to 9769)		
TITLE	Leadlay,P.F.		
JOURNAL	Direct Submision		
PUBMED	Submitted (24-SEP-1990) Leadlay P.F., Dept. of Biochemistry,		
REFERENCE	University of Cambridge, Tennis Court Road, Cambridge CB2 1QM, UK		
AUTHORS	Location/Qualifiers		
TITLE	1. 9769		
JOURNAL	/organism="Saccharopolyspora erythraea"		
PUBMED	/mol_type="genomic DNA"		
REFERENCE	/strain="NRRL 2338 (sub-species)." "		
AUTHORS	/db_xref="taxon:1836"		
TITLE	57. 9593		
JOURNAL	/gene="eryA"		
PUBMED	57. 9593		
REFERENCE	/gene="eryA"		
AUTHORS	/codon_start=1		
TITLE	/transl_table=11		
JOURNAL	/product="6-deoxyerythronolide B synthase"		
PUBMED	/protein_id="CAA39583.1"		
REFERENCE	/db_xref="GI:46976"		
AUTHORS	/db_xref="GOA:003133"		
TITLE	/translation="MSGDGMSTPEKRIKTKTVTBLDSVTAKRLFEVHRGEPRLAY		
JOURNAL	/db_xref="UniProt/Swiss-Prot:003133"		
PUBMED	GMAKRFPGVDSPESFWEVFGGGDIIAAPPDRQWEPDPDRRLGGMIAAAGDPDAGP		

mat_peptide

40.3%; Score 242; DB 1; Length 9769;
Beat Local Similarity 64.9%; Pred. No. 3,5e-14;
Matches 409; Conservative 0; Mismatches 185; Indels 36; Gaps 2;

ORIGIN

1 CCGGCGTCCGCGAGGCTTACACCCGGGGGGTGGAGGTCACATCGGCTACCGCAGTGGCG 60
Db CCGGCGTCCGCGAGGCTTACCGCGAGGCTTGAAGGTCACATCGACCCCGCTTGGCGG 2638
QY AGGAGCGCCCGGTCGACTGTCGCGGTCGACCCGTTCCAGACGACGAACTTCTGGCTCCGG 120
Db ACGGCGGGCGCGGTGAGTCTCCGCTTACCCGTTCCAGGCGGACGCTGCCGCTTACCCCT 2757
QY TCCCTCTG-----GGCGGGTCCCGACACCGGCG 150

terminator
promoter
misc_feature
gene
RBS
CDS

GAAPSALAGLDALPEVPAATEREELVORLERMIALALPVAQADASTGANPSCD
DLGEAGVDELLEAGRELDDGNSHHHHH"
10815..10861
/note="T7 terminator"
10882..10899
/note="T7 promoter"
10902..10926
/note="lac operator"
10933..21677
/gene="eryAII (morphed)"
10933..10938
/gene="eryAII (morphed)"
10947..21677
/gene="eryAII (morphed)"
/note="similar to Saccharopolyspora erythraea EryAII of
Genbank Accession Number AF61566; contains His-tag"
/codon_start=1
/transl_table=11
/product="EryAII (morphed)"
/protein_id="AAV39551.1"
/db_xref="GI:54778586"
/translation="MTDSKVAEYLRATITDLPAAQRIRIELSDPAIYSMACRLPG
GVNTPQRLMBLREGSTLSGFTDRGMILARLHHDPNDPNTGSYDKGFLDDAAGF
DAEFVGVSPREAAAMPQORLLLETSMELNAGIDIPHSIGTATGVLGAVKRGGE
DTAAEDVEGYSYTGAPAVASGRI SYTMGLESPISVDPA CSSLVALHAYELAK
GESSMAVGGAAVWATPGVDFPSROBALADGRSKAFGAGGFGFSEGVTLVLER
LSRRRNHGEVLAIVGSAALNODGASNGISAPSGPQORVIRQALSGCEPDDVAV
BAAGTALADPIBANALLDTGDRDADPPLWGSYKSIIGTQAAQTGLIKVLA
ALNRGELPATLHBEPTPHVMSGGVALLAGNOPMWRGRTRRAASAGISGTNAH
VIVEBAPEHREHTTHDGRVPLVVSABSTALRQAQIAELIEPDLADGIVG
LATTRAHERRAAVASTREAVAGLEIAAGATADAVEGTEVGRVAVFLPFG
GSMWAGKAGELSSPPVPAKIRACDSMA PMODKVSVDLRQAPGRLDRDVOP
VLEAVWVAALIEWRSGVPEPAVVGHSQSIIAAHVAGALTIEDNALTVGRRLKAS
LSEGGMAVALGBAAVERKLEPMQDLSVAAVNGPSSVSGPGALRAFSDCAE
GIRVDIDVDYASHSPQIERERBELLETTDIA RPARVTFHSIVESRMDGCLAE
LYRNRLRETVFADAVTRLAESGDYAFI EYSPHVVVAVEBAVEADGADAVVGS
LHRDGLDSLAFSMATAHVSGVDIRMDVALPGAAPALPTYPEFORIKLOAPAA
ASDELAVRVSMTPIEKRESGNDGMDLVVTPILSPETWMLCEAIINRGRLACRD
TSASRTEMAQVAVAGTGRPGVLSLSSDSACRPGVPACAVGLITIVOLGACVDA
PVNCTOGAVRTPADDDIARPAQTAAHGFQVAGLELPGWAGVVDIPESVDALATL
LVAVLRGGRAEHLVARDGRHGRVRVASLPQSGSRWTPHGTVALTCBAEPVGO
LVWMLADRGARLVLAGACPDLLAAVEBAGASVAVCAQDAAVRELBALDEPTALV
HAGTLNFGSISIVAPEEFAETIAKALLAVLDEVLDGAVREVEYCSVAGWGA
GMAVVAAGSAYLDALAEHRAHGRSGSVATPMLBGA VDGXYLERELSLDR
AKRTERTVLAAGVAVADVDMPVTSBGPATRPFLPHELGRGQABAEPSGPT
GEPAQRAGISLSPDOENLLELVANAAYEVLGHSAAEINVRASLGSLSINAMAL
RKLSASTGRLPASLVFHPHTVALAQHRLARLSADDAARVAGAADESEPIAV
GICRPFGGISGPEOLMRVLAEGANLTTPGPARGMDIGLHYDPNDPNTGSYVDKG
FLTDADPDGPGFITPRBALAMPORMLETRAWEVERAGIDPDALRGTDGVPVG
MNGOSTYQMLAGABERYDGYQIGNSASVLSGRIATTFMGBALTYDTRCSSLVGI
HLMQALRRECSLALAGVTVMSDPTTFVDFSTQRLAIDGCKASADAPALSE
GVALVLELSPRANGHQLAVLRGSAVQDASNGLAANPNSQBRVTRQALASG
VPADVDVVAHGTGTELDPIBAGALIMATYGDRLPLGKVTNIGRTQAAAG
GVTKVTLAMHGMGLPSLHDELSPHIDELSGAVELRESEVPAGNGIRPAGVSSG
VSGTNHVTVERAPARQEAATRGRLPVLGSRSEAVVAQARALAEHRLDPELGT
TDAATTLATGRAPFVRAAVLGDPRACVCAELALDEGRSALAVPITSAPKRPVAV
PFGQAQWGWEMARDLLESSSEVFAESWSCRCEALSPHMDLADVNGDGPDEHARD
VLOPVLSIMVSLAEIMRAHGTPAVAVGSGEIRIAAHVAGALSLEAAKVVALSV
VIRELDDQGMVSGARDELFTVLARMDRVAVAANGTGTGTVAGLELDEFPAB
AABEMKPRRIAVRVAHSPVARIBERLALGTTTAVGSPVLSHSTVAGEYDILV
MBSVYRNLRPVLPEOAVRGVVEGSPDTPVSVSPHVLMAVERBAHGADELTCV
PTIRREGOSHEFTLRILRAHVGQADLRPAVAGRPALPTYPPEHQFWRPRPR
ADVSAVGSRGAHRLILAAVDVFGHGAVALGSLSDQEWLAEHVAGGCTLVPSVL
VDLALAGBDVGLVLEBELVLRPLVLAAGALLRMSVGA PDSGRTTIDVAAEDVA
DLADQMSQHAQTLAGVAGAGPDEQMPEDAVRIPLDDHDTGLAGLGEQYEGSPFO
ALBAARWKDSVVAEVSIADEBEGVAFHPVLDAVQOTLSIALGEEGCGKLPFANNT
VTLHAGSATSVRVATPAGADAMALVTPDAGHVLATVTSVLVRSCEKKEQDEPRGG
EGEILHDMVRLABPGSTGRVVAALASDDDAVTRSGEPEDAVLVRYBEGDDPRAA
RHGVLMALVRLMLEQEBLPGATLVLAISGAVTSDDSVPEPGAAMGVICRQA
ESDPRLVLTDDAPGMLPAVPNDPQALRGDVPFRLSLPALSMTLPAGTORLVP
GDGADSVLEPAPADVQPLRAGEVRDVATVAFNPDLPLASMLPQKDMGTEAG
VTVAVGPDVDAFAPAGDVGLFGQAFAPLAVTQTHRLIARVDDMSADAAAVIATTT
AHYALHDLAGLRAGQSVLTIAAAGCGMAAVALARAGAEVLTATGPAKGTTRALGT

DEHIASRETFARKFRRTGRGVVDVVLNLSLTGELLESADLLAEDGVFVEMKTD
LRADGPRGRVYAPFDLGEAGDRLGELILEVEVLLAGELDRIPVSMWELLSAPALQ
HMRGVRGKLVLTOPAPVDPDGTVLITGGTGTGLLAEHLVTHBGVRLILVSRGG
ADA PGSDLELRAEIEDGASAEIACDTPABDALSALLDGI.PRLTGVVHAAGVADL
VTSIDPEAVEQVIRAKVADAMNITHEITANTGLSPFVLFSSAASVLAGPQGVTAANE
SINLALALRRTRRLPAKALGWMQASSENTSGLGRIATGVAALEPTEKALAFDSA
LRGGGVVPEPLSINRSALRRAPFVPLRQMVAKIRACQASAGAGVNVDRLAGRVE
SDOVAGLAELVRSAAAVSGYGADOLPERKAFKDLGFDLSLAAYELNRLGTAGVVL
PSTLVDPDPLPLVAEHLDRLSASAPAYDIGRDLREKALSALEDGHDVYGQL
ESLRRMNSRRADAPSTSAISBASDELPSMLDQRFSGGEDLGNSHHHHH"
21742..21788
/note="T7 terminator"
21809..21826
/note="T7 promoter"
21829..21853
/note="lac operator"
21860..31407
/gene="eryAII (morphed)"

Query Match 31.5%; Score 189; DB 11; Length 31531;
Best Local Similarity 59.5%; Pred. No. 1.6e-09;
Matches 359; Conservative 0; Mismatches 235; Indels 9; Gaps 2;

Qy 1 CCGCGCTCGCCGAGCGGTACACCCGGGGGTGAGGTGACCTGGCTACCGCAGTGGTG 60
Db 24455 CCGCATTTAGCGGAAGCCTATGCCAGGGCGTGAAGTGTGATGCTACCTGTTTCGG 24514

Qy 61 AGGAGCGCCCGGTGACCTGCGCGGTCTACCGGTTCAAGACAGAACTTGTGCTCCCG 120
Db 24515 ATCCCGCCGAGGTGAATTAACGATGATCCGTTTACGCTCAAGCTTACGCTGCGA 24574

Qy 121 TCC---CCCTGGGCGCGGTCCCGGACACCGGCGACAGTGGCGTTACCACTGCGCCG 177
Db 24575 TTCCGAGGTGGCGGGCTCTGTCAGCAAAATATATATGTCGTTACAGCTGTTGGC 24634

Qy 178 ACCCGTCGACCTCGGGCGGTCTCCCTGGCGCGAGCGGTCCTGTGTGACCGAGCGG 237
Db 24635 GTMAAGCGAATGGGAGTCTGCGTCCCTGCGCGGTGCGTGTGTAACCGGCCGG 24694

Qy 238 CAGTACCCCGGCTGCGACGAGCGTGGTCCGACGCGCTGGAACAGCGCGGCGACCG 297
Db 24695 GTGATCATGTGAGCTGTGCGATGCCATCCGATCAAGGCTGAGACAGTGGGGCAACGG 24754

Qy 298 TCGTGTGTGACACCGGCACTCGCGCGCGGATCGGCGCGGCACTGACCGCGTGCAG 357
Db 24755 TTTTGAATGACGACGTCGAAAGCGGTTTCCACATTCGACAGGCTTGGAAAGCTGTGATA 24814

Qy 358 GCAACGCGCTGTCACATGTGTCTCTGTGCTGCGCGTCCGACGAGGGCGGTGTGACAG 417
Db 24815 CTGATGCGCTGACACCGTATGATCGCTGTAAACCGGTG-----ATGCGAAGCTGTG 24868

Qy 418 ACCCGAGCTGACACCTCGCTGTGTGTCACAGCGCTCGGCGACGCGGATCGACGTCC 477
Db 24869 ATCCGAGTCTGATGCTGTGCGCTTTGTGTGAGGCCCTTAAGGTGCTGTGGGTGGAAGCAC 24928

Qy 478 CCTGTGTGCTGTGACCAAGGACCGCGCGCGCTGACCGTTCGAGACAGACGTGATCCGG 537
Db 24929 CGCTGTGTGCTGTGACCGCGTAAAGCTGTGTCAGGTTTCTGATGTGTGAGTGTGATCTTG 24988

Qy 538 CCCAGGCAATGTGCGTGGGCTCGGCGCGGGTGTGGGCGTGAAGTCCCCCGCGCGGTGG 597
Db 24989 CCCAAGCATGTGTGGCGGGCTGGGCGCGGTCTGTGTATCGAACACCGGGTGTGGG 25048

Qy 598 GTG 600
Db 25049 GCG 25051

RESULT 10
AB088224/c 210614 bp DNA linear BCT 11-JUN-2003
LOCUS AB088224
DEFINITION Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.
ACCESSION AB088224

AB088224.1 GI:30698345

KEYWORDS
SOURCE ORGANISM
REFERENCE AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE AUTHORS
COMMENT

Streptomyces rochei
Streptomyces rochei
Bacteria; Actinobacteriae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

Kinashi H., Fujii S., Hatani A., Kuokawa T. and Shikawa H.
Physical mapping of the linear plasmid pSLA2-L and localization of
the eryAI and actI homologs
Biosci. Biotech. Biochem. 62, 1892-1897 (1998)

Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H.,
Naito, O. and Kinashi, H.
Identification of two polyketide synthase gene clusters on the
linear plasmid pSLA2-L in Streptomyces rochei
Gene 246 (1-2), 123-131 (2000)
10767533

Hirateu, K., Mochizuki, S. and Kinashi, H.
Cloning and analysis of the replication origin and the telomeres of
the large linear plasmid pSLA2-L in Streptomyces rochei
Mol. Gen. Genet. 263 (6), 1015-1021 (2000)
10954087

Mochizuki, S., Hirateu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K.
and Kinashi, H.
The large linear plasmid pSLA2-L of Streptomyces rochei has an
unusually condensed gene organization for secondary metabolism
Mol. Microbiol. 48 (6), 1501-1510 (2003)
12791134

(bases 1 to 210614)
Mochizuki, S., Hirateu, K. and Kinashi, H.
Direct Submision
Submitted (15-JUL-2002) Haruyasu Kinashi, Hiroshima University,
Department of Molecular Biotechnology, Graduate School of Advanced
Sciences of Matter, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima
739-8530, Japan (E-mail: kinashi@hiroshima-u.ac.jp,
Tel.:81-824-24-7869, Fax:81-824-24-7869)
The nucleotide sequence has been determined by using restriction
fragments and nested deletion fragments of the ordered cosmid
library of pSLA2-L.
pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries
143 ORFs.
Gene prediction was based on the unique codon usage in Streptomyces
(Bibb et al., Gene 30:157-66 (1984)) using the Frameplot program of
Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as
implemented at
<http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>. Where possible we
chose an initiation codon (atg, gtc, ctg or tgg) which is preceded
by an upstream ribosome binding site sequence (optimally 5-13 bp
before the initiation codon). If this could not be identified we
chose the most upstream initiation codon.
Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes
(three PKS gene clusters for lankacidin, lankamycin and an unknown
type II polyketide, a carotenoid biosynthetic gene cluster, many
regulatory genes and others). The range of each biosynthetic gene
cluster has been deduced by comparison with similar gene clusters
in most cases and may be revised in future.

location/Qualifiers
1..210614
 /organism="Streptomyces rochei"
 /mol_type="genomic DNA"
 /strain="743AN4"
 /db_xref="taxon:1928"
 /plasmid="pSLA2-L"
 /note="linear plasmid"
1..1992
 /note="left terminal inverted repeat, TIR-L; shows 99.4 %
 (1981/1992) sequence identity to TIR-R (complement
 [208623..210614])"

CDS
 /note="N-terminal sequence is almost identical (435/437)

with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L
ORF1 (501 aa)
similar to AL590463 Streptomyces coelicolor putative helicase, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L"
/codon_start=1
/transl_table=1
/product="putative helicase"
/protein_id="BAC76459.1"
/db_xref="GI:30698346"
/translation="MSTSTRTDQRRAAQAEVADVAVALEIPVLLAPERGLRQIVIMATGSKTRVARSSEKLRAGVLVPLSDLTLTTEAAMEAGTGMIGVSSLRGSDVAFPTDVEELIYNVRPEPKVYVPAVYSLGQTLERARQGLPGMDLVNDEAHTSGRLKPAVAVHNDTRIPSLRLYMTATPRIMQJLDEBEGAPGELVASMEDDPDGLFGARCFLLTISEA1IDRGI CAPYQVVCVDITDQLQAQLLGVGRDEVRGARLALQALILKASSENFEFRTLVFHHMVKEAAFAAGIPDVAKRILHAAGPGIYPTIWMNMLCGHKKPHRRRLRDIPTGIDTGVVKEGFLGSKVYKGEVDRECDSPVYMAVDRGSMPLVOAVRARLIDPGGKVASIVLVPLTDPGETADMMLNSRPYNGLLALFRAQPLITGHGEEGRPAAPFPEFVRPVLVGVGAQRGASAGPANSITSPSAMSSISMRAP"
complement (2315...3595)
/note="ORF2 (426 aa)
similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)"
/codon_start=1
/transl_table=1
/product="putative secreted protein"
/protein_id="BAC76460.1"
/db_xref="GI:30698347"
/translation="MRKKRPALIGIPGLVAVLAPAPAAHSAADTGVGTMAAPTA PAISTVEDDTLQCTVHLSVAGSLVRLNREKSTPLTIGTHARPAAGQATVADPGTDVVRFGGRSATLAPGTQKMSPPVALPTTGQDVIYSLTLPRTGGSTYHSAIYQHNVAAGDVTGAADLTLPVSTATSMHFLSGVSYDRRAGTADSSVVTLSGSDTGGTLDADRBRVDLGGARVAVTLGNDVLQPGIAPASDEVTAEELIAGROLIERHLSAKIKYCATITPFGAGITVLTGFTPRREAVRQVNDVMTSGAFTVLDPAVILADPARPDHILPAIDGGDGLHPNDAGMAAARAFFPDSLR"
complement (3659...4645)
/note="ORF3 (328 aa)
similar to Y00459-2 Streptomyces griseus regulatory protein, Strr (350 aa)"
/codon_start=1
/transl_table=1
/product="Strr-like regulatory protein"
/protein_id="BAC76461.1"
/db_xref="GI:30698348"
/translation="MTALAGNLDOETRSIPISLVPGSGPRLCGSDAEHVAALADTD BALPIVERSTSLVYIDGMHRYLAAKAGHTTIVRLFDGAASEAPLAVASNNHTGLPLSRDRRAAQRILAQWPHSPDAVAGIAGTAKYUALRPLAAGTPNPAQRGDDGIRPLDGTGRKAELTAAORPQASVREVARHAGISPATSDYRRRLAAGSEVPERVTPGARPGVSGSRATPGGGAAPSPVSVIRPPEDSPVLRLDRLLHKSGRLLRLTLQCAVERTALLAMATVPPICTDLVAELAREVADLMAEFARBEVARTDG"
complement (4922...5812)
/note="ORF4 (396 aa); lankacidin biosynthesis protein similar to AE004625-4 Pseudomonas aeruginosa pyroloquinoline quinone biosynthesis protein B (304 aa)"
/codon_start=1
/transl_table=1
/product="pyroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76462.1"
/db_xref="GI:30698349"
/translation="MILLGTAAAGGFPQWNCALCARGRRGELPARSQRCAVSGDGRDWMVLNNSPDIRTQLAALPALTPGQGRDTPVGVLLTAAVEHALGLVLRATQTLTYAAPRGALSAELPVRLGLDVAAPMDMVAAPGFAVAGLVTVAHPVGTAKARVAHADPPAPVCAACRIEDPATGALVYAPCLATPMPGDFDLASATCALDGTFFSAGELGRTITSAGAGSIMGHLPVYAGPGGSLAALAHRLRRLTYTHANTNTPLDSSAAHAAVEAGVEVLPDSSELVLT"
complement (5815...6909)
/note="ORF5 (364 aa); lankacidin biosynthesis protein similar to AJ271117-5 Gluconobacter oxydans putative pqqe protein (359 aa)"
/codon_start=1
/transl_table=1

/product="pyrroloquinoline quinone biosynthesis protein E"
/protein_id="BAC76463.1"
/db_xref="GI:30698350"
/translation="MADPAGAPAGMILIEIHRCPICSPNPIEYRREALTCED
WPIILTOARELGIVQMHFSGSEPLARDPDLVGHARLALAYNLTSGVLTAEBAH
DLARGVDHQLSLQDADPAAGALAGARVATLGAARLTAAGPLVATVILHGN
IDTGRVVDLAVDLGADRIELANTOYGMGLRNRLALMPAALAAAREVRAATRY
AGPELVYVADYDPRPKFCDMGSTGLTTPAGVLPSPAAVAILTTLPEVNALRY
PLSEIYASRFAVAGTGMWRBPCRCRSHADHGCRQAQLTCDAAATPDACGL
SPHRSVDAALAVDGPVPAFVPRGVPVA"
complement(6502..7174)
/note="ORF6 (90 aa), lankacidin biosynthesis protein
similar to AL603642-200 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein D (98 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translation="MTGLPEPTVPLRPGVRLTRDPARGELALLPERVAVLNDTAAV
LAHCDGTSLAGIYERLAREYBGSVADRELLRLAQRVVDLHG"
complement(7171..7899)
/note="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translation="MSMSVYREVAAPMSEAFORLHALESSYWDHPRHMEGL
DEGELTMANRWYORCLPQDAIVANPLBVRWMLSRVYHDAACAGAEK
MLIAVGLRDEVDRELTLAGTPADAYDPRARRPMLEAASGTELESPELT
AARLGLRHHYPMIAESGFYFARLVVGPBSRLDLVARHVSREOEACRALA
FKCVLNAVLDSIDYHTGNGATRS"
complement(7926..8030)
/note="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="MRTSGKELPAKAMRHPDFTIDTGMETVAIYPSR"
8420..10057
/note="ORF9 (545 aa), possible lankacidin resistance
protein
similar to AL133424-19 Streptomyces coelicolor putative
ABC transporter ATP-binding protein (544 aa)"
/codon_start=1

CDS

Query Match 22.9%; Score 137.6; DB 1; Length 210614;
Best Local Similarity 56.5%; Pred. No. 4.3e-05;
Matches 366; Conservative 0; Mismatches 234; Indels 48; Gaps 4;

QY 1 CCGCGCTCCGCGGCTACACCCGGGGGTGAGGTCACTGCGCTACCGCAGTGGG 60
DB 61498 CCGCGCTCCGCGGCGGCGCATGTGCGGGGTCCAGGTCACTGGAACCCCGCTTCCG 61439
QY 61 AAGGAGACCGCGGCTGACCTGCGGCTACCCGTTTCAAGACAGAACTTGTGCTCCG 120
DB 61438 CCGCGCGCTGACCTGCTTACCCCTTCAAGACAGCGCCGCTTGGCAACCG 61379
QY 121 TCCCGCTGGGCGGGTCCCGACACCGGACAGAGTGGGCTTACGACTCGCTGGACCC 180
DB 61378 CCGCGCGCGCGCGCGGCGGACCGCGGCGGGAAGCGGCTACGAGGTGGTGGCA 61319
QY 181 CCGTGAACCTCGGCGGCTCTCCCTGGCGGACGAGGTCTGTGGTGGACCGAGCGGCA 240
DB 61318 AGCGCGGAGTGGGCGGCGCGCTCAACCGCGCGCTCTGTGGTGGACCGCGCGG 61259
QY 241 TACCCCGGCGTGAACGAGTGTCCGCGACGCGCTGGAACAGCGCGGCGGACCGT-- 298

DB 61258 TGCAGACGCGCTGTGACAGACACGCTCCGCGGCGCTCCAGACAGCGGCGCACGCTCG 61199
QY 299 ----CGTGTGTGACACCGCGCACTCGCGGCGCGCGGATCGCGCGCACT--CGACCGCG 351
DB 61198 GCACCGCTGACCTCGACCGACCGGACGCTGCGCGGCGCGCGCTCGCGCGCGCA 61139
QY 352 TCGACCGGACCGCGCGCTGACAGTGTGTCTCTGCTCGCGGCTCGCGAGGCGGCTCG 411
DB 61138 CGGCGCGCGGGAACCGGCGGCGGTGTCTGTGCTCGCTGACATGCGAGGGAAT 61079
QY 412 TCGACGAC-----CCAGCGCTGACACCC 435
DB 61078 CCGCGACAGACACCGCGCGGAGTCCGCGCGCGCGCGGACCTGCGCGGACCC 61019
QY 436 TCGCGTGTGTCAGGCGCTCGCGCGACCGCGGATGACGTCCCTGTGCTGTGACCA 495
DB 61018 TGAAGCTGATGACGCGCTGCGGCGACCGGAGTGTGAGGCGCGCTGTGCTGACCC 60959
QY 496 GGGACGCGCGCGCGCGTGAACCTGTGAGAC--GACGTGATCCGCGCGAGGCAATGTCG 552
DB 60958 GCGCGCGGTGCGCTGACCGGACAGACAGGAGCGGAGCGGACCCCGCGCAGACGCGGTG 60899
QY 553 GTGGGCTCGGCGGCGGTGTGAGGCGTGAAGTCCCGCGCGGCGGAGG 600
DB 60898 CCGCGCTCGGCGGCGCGCGCGCTGAAACACCCCGCGGCTGGGCG 60851

RESULT 11
AX697993 4725 bp DNA linear PAT 02-APR-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
MICROMONOSPOREA carbonacea
MICROMONOSPOREA carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1
AUTHORS
TITLE
JOURNAL
FAPUBS
source
1..4725
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"

ORIGIN

Query Match 20.7%; Score 124; DB 6; Length 4725;
Best Local Similarity 55.0%; Pred. No. 0.0029;
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;

QY 1 CCGCGCTCCGCGGCTACACCCGGGGGTGAGGTGACT-----GCGTACCGCAG 54
DB 2642 CCGCGCGCGCGGAGGCTTCAACCGAGGGGTGCGGTGACTGCCGCGGCTCTGGCGG 2701
QY 55 TGGGTGAGGAGGAGCGCGCGTCACTGCGCGTCAACCGTTCAAGACAGAACTTGTGCG 114
DB 2702 GAGCGCGCGCGTCCGCGTGAACCTGCGACGTAACGCTTCAAGCGGCAACGCTACTAGC 2761
QY 115 TCCCGGTCCCTCGGCGCGGTCCCGCACA---CGGCGACAGTGGCGTTACAGCTCG 171
DB 2762 TGAAGCGCGCGCGGAGGAGCTCGCGCGCGGCGGAGACGCTGACCGGTG 2821
QY 172 CCGGACACCGCGTCACTCGGCGGCTCTCCCTGGCGGACGAGGTCTGTGTGACCG 231
DB 2822 GCTGCGGCGGCTTACGAGCAACCGCGCGCGCGCGGCGGACGCGGTGCTGTGTGACCG 2881
QY 232 GAGCGGAGTACCCCGGCGTGTGACGAGTGTCCGCGACGCGGCTGGAACAGCGCGGCG 291

Db	2882	GC	CCG	GAG	CAC	CCG	CG	AGG	CTG	ATC	GAG	GGG	CGT	AGC	CGC	TCA	CCG	CG	CGG	CG	2941
QY	292	CG	ACC	GTG	TGT	TTC	GCA	CCG	CG	CAG	TGG	-----	CG	CGC	CGG	ATC	TGG	CG	CGC	339	
Db	2942	CCG	AGG	TG	CGC	CTG	TGA	CCG	ATC	GAG	CGG	ACA	CAG	CA	CGA	CCG	ACG	CGG	CGG	CGT	3001
QY	340	CAC	T	G	A	C	G	A	G	C	G	G	T	G	A	C	T	G	A	T	390
Db	3002	TG	CT	C	A	C	G	G	C	G	G	T	G	A	C	C	G	G	A	C	3061
QY	391	-----	CG	T	G	C	G	C	C	G	A	G	G	C	G	T	G	T	C	A	444
Db	3062	GT	CG	CG	CG	CA	CC	CG	CA	CC	CG	CG	GT	CG	CG	CG	CG	CG	CG	CG	3121
QY	445	T	C	A	G	A	G	C	T	G	A	C	C	G	A	T	C	C	C	T	504
Db	3122	CC	A	G	A	G	C	G	T	G	A	C	C	C	T	G	C	G	A	C	3181
QY	505	C	G	C	C	G	T	A	C	G	T	G	A	C	C	A	T	C	G	A	564
Db	3182	T	C	T	C	G	T	G	A	C	C	G	A	C	C	G	A	C	C	G	3241
QY	565	G	A	G	T	G	A	G	C	T	G	A	G	T	C	C	C	C	G	A	600
Db	3242	G	A	G	T	G	A	G	C	T	G	A	G	T	C	C	C	C	G	A	3277

RESULT 12			
AX697977			
LOCUS	AX697977	60196 bp	DNA
DEFINITION	Sequence 1 from Patent WO03010193.		linear
			PAT 02-APR-2003

ACCESSION	AX697977
VERSION	AX697977.1
	GI:29499034

SOURCE ORGANISM	Micromonospora carbonacea	Micromonospora carbonacea
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE
AUTHORS
1
Farnet, C.M., Staffa, A. and Yang, X.

TITLE
JOURNAL
Genes and proteins for the biosynthe
Patent: WO 03010193-A 1 06-FEB-2003;

Ecopia Biosciences Inc. (CA)
Location/Qualifiers

```

1. .00196
Source
/organism="Micromonospora carbonacea"
/mol_type="unsequenced DNA"

```

```

/inv_type=unassigned DNA
/db_xref="taxon:47853"

```

Query Match 20.7% Score 124 DB 6 Length 60196

Best Local Similarity 55.0%; Pred. No. 0.0012;
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4

1 CCACGCTCCGCCGAGCGCTACACCCGCGGGGTTGAGAGTTCGACT-----GGCGTACCGCAG 54

Db

38933 CCTCGCGGCGCAAGGCGTTACCCGAAGGAGTGGCCGTGAGACTGGTCCCGGCTTGGCCG 389

OY 55 TGGGTGAGGSA CGCCCGGTGCACTGCGGTACCCGTTCAACGACAGAACTTCTGCG 114

Db
38993 GGGCCGCGGCGTCCGCGGTGACCTGCCACGTACGGCTTCCAGCGGCAAGCTAATGTC 39011

115 TCCCGGTCCCCCTGGGCGGGTCCCCGACA---CCGGCAGCAGTGGCCGTTACCAAGCTCG 171

Db 39053 TGGAGCCCGCGCGAGGCGGACTCCGGCCCCGGGCGCGGGACGGGCTGGCGCTACCGGGTGG 3911

172 CTTGGCACCCTGTCGACTCGGGCGGTCTCCCTGGCCGGACGGGTCTTGTTGATACCG 231

Db 39113 GCTGGCGGCGCTTCAAGCGCACCGGCGCGCGCGGACCGGATGGCTGGTACCG 391

232 GAGCGCAGTACCCCGGCTGGACGACGTGTCGCGACGCGCTGGAACAGCGCGGG 291

Db 39173 GCCCGAGCAGCCGGCGGAGCTGGTCAAGGCGGTGCGGACGCGCTTACCCGCGCGGGGCG 392

OY	292	CGACCGCTCGTGTGTGACACCGGCAATCG-----CGCGCCGAATCGCGCGCG	339
Db	39233	CCGAGGCGCGCCGTGTGACCTGTGACGCCACGACACCGGACCGGCGCGGTGCGCGGCGT	392392
OY	340	CACTCGACCGCGCTGTGACACCGGACCGGCGCTGTGTGACATGTGTCTCTGTCTCG-----	390
Db	39293	TGTCTCACCGGCGCGCGGTGCGCGGCGCGGCGGCGACCCCGGCGTGTGTGTGCTGTGCGACCGAATC	39352
OY	391	-----CGCTGCGCGAGGCGCGGTGTGTGACGACCCCGACCTGTGACACCTTCGCGTTG	444
Db	39353	GTGCGCGCGGACCCCGACACCGCGCGCTGTGTCGTGCGCGCGCGCGCGGACGTGTGCTGTGA	39412
OY	445	TTCGAGCGCTTCGCGCGACGCGCGGGAATGACGCTCCCCCTGTGTGTGTGTGACCAAGGACGCGCG	504
Db	39413	CCCGAGGCGCTGTGCGCGACCGCTTCGCGCGCGCGCGCGCGCGCTGTGTGAGTGTGTGACCCCGAGGCGCG	39472
OY	505	CCGCGCGGACCGTGTGCGGACGACGATGATCCCGGCGCGACGCGCATGTGTGCGTGTGAGCTCGCGC	564
Db	39473	TCTTCGTCGCGCGCGCGCGGAGACCGCGCGACGACGCGCCAGCGCGCAGGTCTGTGGGAGTTTCGCGC	39532
OY	565	GGGTGTGTGCGCGTGTGAGATCCCGCGCGCGCGCGGTGCGGTG	600
Db	39533	GGGTTCGCGCGCGCTTCGACATCGCGCGCGCGACGTCGCGCGCG	39568

RESULT 13					
AF263912					
LOCUS	123580 bp	DNA	linear	BCT 24-MAY-2000	
DEFINITION	Streptomyces nourisel ATCC 11455 nystatin biosynthetic gene cluster, complete sequence.				
ACCESSION	AF263912				
VERSION	AF263912.1	GI:8050835			

KEYWORDS	
SOURCE	Streptomyces noursei ATCC 11455
ORGANISM	Streptomyces noursei ATCC 11455

Bacteria; Actinobacteria; Actinobacteriidae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
1 (Dates 1 to 123580)
AUTHORS
Brautaaset, T., Sekurova, O. N., Slettta, H., Ellingen, T. E., Serlm, A. R.,
Witt, C. and Zetshen, S. D.

TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in *Saccharomyces novaezei* ATCC 11455: analysis of the gene cluster and Valia, S. and Zolnerov, S.B.

JOURNAL OF
Chem Biol 7 (6) 395-403 (2000)

PUBMED
CURRENT CONTENT: 1 (0/1) 225 100 (100%)
10873841
REFERENCE 2 (bases 1 to 123580)

AUTHORS
Brutaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R.,
Valla, S. and Zotchev, S.B.

TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim

FEATURES

N-7489, Norway

Location/Qualifiers

```
source
1. .123580
/organism="Streptomyces noursei ATCC 11455"
```

```
/mol_type="genomic DNA"
/strain="ATCC 1145"
```

```
/db_xref="ATCC:11455"
/db_xref="Eaxon:316284"

```

gene complement (46. .783)
/gene="ny8p"
complement (46. .783)

```

CDS
    complement(10. . /63)
    /gene="nybP"
    /function="tryptophan biosynthesis ACP

```

```
/transcription_probability_pos_domains=1 "transcription_pos_domains_on_PKS"  
/note="mutative 4'-phosphonothreonine transferase"
```

```

/next generate :
/codon_start=1
/translate=1

```

```
/product="NysP"  
/protein_id="AAF71762.1"
```

```

/db_xref="GI:8050836"
/translation="MIEILLPATVATEAYDDRRPGDRLLSSSESVIARAVESRRRE

```

FTTVRHILARRALRLGLHDDRAILPNRRGAPQMPPGIVGSMTHCAGYRAAIVSPALSLA

gene
CDS
/gene="ny8g"
/complement(867. .2684)
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AA071763.1"
/db_xref="GI:8050837"
/translation="MSPDDISEERTAPRPARIVGLRPHRSVALAVSMGVGIVL
NAFGPILGRVYDLADGVLCGYPGAPGIDPAITRLLVTLALVYASLPMLAGR
LVASAVRTIHRLRDAREKTRPLPHRDROPAGELISRTINDINLOQTQTLAE
LITSLISLTMVLVLMVISPSLAVMLSVPSALIAARI SKRAQPHVAAVMFANTL
NAHBEVCTGHALIKGFDRRAAEERFDACNDVYRAAKAQPASQPMVNFANTL
GVLVAVIGAMVINGTITLGDVQAFILYAROSPIVEIASVAGSLQSGIASORVF
TLDDAPEOAPDLRPGTPARAGRVETSPSPDPLINLSLTVPGSTVAIVG
PTGAKTTIGNLMRYEPDSEGRILLDGTDTMTTRDILRSRGLVQDTMLFGGTTA
ENIATGAPGACRADIIEAARATCADRFITLPGQYDTVLDDSGTVSAGEKQLITVAR
AFLARPVAVLDEATSSVDTRTEVLIQRMNSLRAGTSFVIAHRLSTTRDAILVVM
DAGRIVEQCTHQLCAQGLVYARLHAARTHTPTAGAAG"
/complement(2662. .4416)
/gene="ny8h"
complement(2662. .4416)
/gene="ny8h"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AA071764.1"
/db_xref="GI:8050838"
/translation="MLRLRAQLRPYAWATAALVALQVLIIGTLLPTLGAALIDQ
GVVGGGRITELGVVGVVALVOIAALAAALAAATATMGRDLSALPRITLDS
AEIRGCRPSLITRSVNDVOVONLAQYRGIVCAPIIMCAGSITLAIROVPILAL
LVALVAVAVCGLLAARNGITYARQTLTDLRGLRCAITGVRSFVDDHERA
RFAQNDALVVSRRVGLIATMLPVLLANNGPTVALMTQSHRIDARMPIGISIA
LISYLSLIMSVMVLA FVL SVPARVACGR IAEVLDTGSSVAPPAQPVPGPARI
ELCAGYRPGAEPEVLRDVLVEBEREIAVIGSTGSKTLLNLVPLVRLATGSAV
RYGGTDVRELTAATLAAVGFYRORPYLPSGTVANLRCPDADTDEBELARVACV
ADPVARMPDGLDAETIGCGANVSGGRORLSLARALLRPSITLDDCCSALDQITDA
ALRTALVPTAGATVITVAKRISAGDADRIVLDRGVVAGCTHVLRTSPYREI
ALRSQTEBAAGLGRP"
4714. .5748
/gene="ny8DIII"
4714. .5748
/gene="ny8DIII"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dcbp-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="ny8DIII"
/protein_id="AA071765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGODGSYLAEHLSQGYOVWGLIRGOANPRKSRV
SLIASLELPIIDGLMDQSLVAVDTVODEVYNIGALS FVPMWQOALVTEVANGV
VLRMLELRIMVGLSTSRVSPRGQIRFYQASSSEMFGAETPORBETTLPHRSPYG
AKAYGHTTRRYRSPGVAVSGMLFNHRSRGQOERPTKISLAVARIKQGLDGL
ALGNLDAYRDMKQADYRYAAMHLMQODAGDYGVTGQGMVRVADVRIAEHVEGAM
EDYVIDPDLVRPAEVEVLCDSATAQDLKPKVDVPTLKRMMVDSDLQAVSRENO
YGDVILAAAMW"
5930. .34363
/gene="ny8I"
5930. .34363
/gene="ny8I"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1

/transl_table=11
/product="NysI"
/protein_id="AA071766.1"
/db_xref="GI:8050840"
/translation="WDNEQKLDYKLATADILRRTTRRVHKLBSAAQBEPAVITGMTR
YFGVRSPEDLRMWVAGEHGTVPPTTRGMDELAALAPTSGGTFLADAPPDADPF
GISPREAVAMDPOQRVLSLEAEEAFEGATIDTSYSGSTGVFI GMAQDYRGADPF
AGSPOLTGNTGVSLSGRISYTFGTGAPATVDTACSSSLVAHIALTOALRACETIAL
AGVITLMSGRCFIRMGQCGLSADGRCSREDTADGTCMAAGVILVIERLSAYRN
GHEITLAVNRGINNDGASNGITAFRNSQOVITQOALVANRLACDIDVVEAHGTT
TLGDVPEAQQALLATYGQNRPADRPLLSVKSNIHVTOAAAGVAGITKNVAMRHGTL
PRTTHAEPTTHHVDSOGAVRLITDTVTWVPPVLPAGARAAVSFGISGTAAHTIIBQAP
EPOPDAAIADDDAAGSTPATVPTVVGVVPLSGRTPALSGOAAALAAALDTGRRP
DILDLASLIATTRAGFEHRAVILATDHPALTDLTALDADDDPAAAPATVCTTAAET
RLAVI PTGCGAORLCAGBELARPPATDIALDALDPTPHLDRLPRLKCTDALL
DRTAYAQPLPAVEVALRYLIESFGVRPDHLGSHVGI VAHLGVLSLADATVVA
AGRMLOALPDGAMIAVQASADVA PLIAGDEDDVAIAVNGPSAVVLSGAEATVTA
LAEQALQGRKTRRLRVSHAFSPIMEPMLDPAFVVEEDTLQPLPVSNLTGKPA
TVAQLSADYVDVDRHARVPADGIDMLABHDTTAFLELGPQVVSAMQDCLDADA
DAVTLPAIRAGREHHTLTTLAGLHNGATLDTMGCPCAGTCARPTDIPYAPORRY
WPKALQSTADLRSGLCAHHPLLSAAVSLADAGCTLITGLSKOTHPMLADHTYRG
TTLTPTGAPLEIAVAGDEVCDRVEBELTAAVPLLPQGGVQVQDMLGNPVSQKRT
VVNARPTGDDTPTWTAATGVLITPADSRQLPASBOGCTPLADPHALDAAQMP
AGABELPLDGHYRLADGFGYGVPGCGLRAAMRGADVVAEVLPEAGRSDAEGL
HPALDALLHAAPFTGLGRRGGGLPFGMEGSLHAGGATTLRVLTPVADALALT
ADGCAVPLSVDSILVRSVATQOULTAAVANDALPRLDPTTPVQAPADRGVALLGA
DPGILLTHAGPADAPATYDILANLAAADPPTVTTVLSLIGTDDAADPARSKACAE
ALAAQVTLMDHHERFAARLVEFTGATVGRVAAAANGVLRSAQSENPQCFALVDL
DPDGAAGAAALVAALVAVSGEPOLAVRGDVLRYRLVRLPTEVGAQADGTGVDGDSG
VSFGSEGAVLVYTGCGIGAVLARLVAEYGRDILLVRSSEGRVAGAGELVAELAG
GARVHEVACDVTDRAAVVELVGHGAVSAVHAAGVLDGMVALLGERLSAVLRPVD
AVMHLHEATRGDLDLDPVYFSLAGVRSRGQANAAAPATDAPTRARAGLPELS
LAKPWEBSGCGMTGTLTVDDERLARSGVPLSVAGLALPFAVAGTDATCVPVRLD
LPLVLRAGVBPPLRLSLRVARRAVAGSAVAGLALQRLRLRDLDBRDEMYLDVRG
QVALVIGHATGADVDRAGARFDLGPDSLTAAVLRRLATVTLRLPATLVFPYVTRH
LATVYDLBELGTDAEVAIVQPAVAVADDPITVYCMACPYGVSSEDLMEVLTGCT
DAVSGEFPNRGMDVDSLXHPDDHPGTSYTGSGGLHABRGEDPQPFMGSPREALTD
SQOQLLESSWEALIRACIDVPSLGSSTGVPAQMYDYDSMLASPEREGQSGSS
PLSAGRVATLGLBGPVAVTDATCSSLVAMHVMQALRSCEGLALAGVTVWSTP
AVFPFAQRIGLSPDGRKAPADADVAGSAGWMLVLEBSDVVRNGHEITAVNGS
SAVNDGASNGITLWNGSQORVIRLOALASGLTGDVVDVAEHSQGTTLGPITEQA
LLATGRSREPERPLPLGSSVKSNIHOTOAAAGVAVIKMVLVAGHGVTRPLVAPAS
SHVMSBEAVELLSROAMPERTGRARRGVSFGISGNVAVIVQAPAKAIAAAGA
ARPTGAVPVILSGGRALKQARILGILQARDAELVVALSLATTRSGEIRQAA
VVAQDRDILASLALADRDPDAVBEBAERGAETVLEQSSQRAAMPSELEHVVQ
PEFAAPFAPVACVDPDLDRPREVPELEDSDEALIDETGTOPALPFAVVALFRL
VESMGVRDPVAGHSIGELIAAHVAVGVLLEDCRVLADRTIMQALPGCAMIATOA
TEDEIAMLDDTVAILAANGPOSVITSGDEBAERTIATTPAERGKTRIRVSHA FHS
PRMGCMALDPRIVAGELTYRARPITPVSDLTGRRDDAVCTAETWYHVRVAPAD
CVRTLRDAGATTFLELSDGLITMAABDTLGDHDAELVPMLRAGARBLAATLAR
LQVRGVDDMAVYLAGTGARRTDLPTVAFQAHYVPOLPTPAALAAADPADQOOLMA
VERGDAERLADILIGIEBODTLPLDSLALTSMRGNOBEKHLIDTLRAREVTRLSKP
TAPVLDGTMVLVASDAPADOPALLDGLADLHSGAVRRLLDGSCARDVLAERL
ARTADVAAATQVLSVLPIDBERDADDCPLTGLALTVLALVQALADTGCAOGRMLVATRG
AVSTRPADPYTHPQAAKWCGRGYALEHPRLMGSLYDLPVYFPERAQRAGLITLAVK
DAPDGEDVALAKATGVSGRUVRTVEALPTPAETATGTVLITGCTGLAEVARKL
ARAGAQLVLTSSRRGPVAGAEARLAELEGPSVSAVCVADADALAAVLTALPER
LPLTGVHTTACVGHYGPDLTSTAEPACTLAKKAGANHMLADRLDPEVLFSGSI
LAGVSGNQSAYGAAANALIDALALHRRARGLAATSANGPAALEAMADNDVETSLR
QGLDGLDAPMTELRARVAVRODVTYVADVDMQYALPTFASARLILAGLPEVAL
AADRTBEDAPAMTELRARVAVRODVTYVADVDMQYALPTFASARLILAGLPEVAL
RDVGSDSITAVBELKRCDAATGSLSPSTWMDYPTPLLELAQYLRABITGAVLEVAGPV
ATGADDDPDLAIIGMAKCFPGGVSPBOLMVASGDTALISFPVNRMWQGHLEDPD
PDRPGTGYTSGFLHAEDEDPDPTFGISPEBALVMDQOQLLETWESPERADIRP
EHLSTLTGTFTGVSQSYOEYGAAGDGEHNVHTSSPSVLSGRSLYVGLSGSPATV
TACSSSLVAILHILACSLANGBSNLVAGAGATIMTPNPIAFSROALVAGSCRAVFS
DDAGMTLAEAGVAVILVERLSDAOGRGPAVAVRGAISLADGASNGLTANGSCQOR
VIRQALNARLAPEDIDLBAHGTCTPLGDIIBQALPATYTGRRDPESALLISVKS
NIGHTQSAAGTAVIKWMLRHSBELPPTLHAPSSHVDSAGVRLTLDARMAPET
GRPRRAVSSFGISGTANAVILEQAPADVATAEERPAVAPVIAAGVWPVWTASAA
ALRQGAERLILHARTVGTALPAAGPLDGLSLVBARAFERAVVVPAGVDPLAALR
AAVADGSPVAVAGVADVEGRVTFVPCGQSOVMGMSGLDDESVPARLAECAAL
AEFTDMSLVDTLRGVGAPSLERVDVQPAFVAVVSLAALMRSGVLPDAVVGHSQC

ETAAVVSGLSLRDGARVVALSRAIGRALAGGMMGVALSVLVEBRLVFEGRV
SUAUNGPSVAVAGPRLALDLHARLTADDLRARIADVAHSHOVBDIHELLELV
LAEALRTSEVPEFSTVTDNMLTAMMDGYRNRIRGRVPRADAYADLLAEYRFLV
EVSHPLSVAEADIDENGVPNAVAGTRDRGGGTRFLSLAEYFVSGVDYDRAGL
FBEQASRIDLPYAFQHEHMLAVPAPRAVAAADPDADLFWAVSDGVSLTALAG
TDEDSVAAVLPLTISWRARARSRSTVDAMRYRAADLGSTLPHSLTGTMLLVTDG
IDTDVAGALETYGAERLVLDEBCVDRLVLERLAGAEVDYGISVLAASERTAV
PGTSLVGTALTVLALIOALGDADIDAPVMTLTVGAVSTGADDELTAPOAOVTGIGMT
AALEHPRQWGTITDLRALADBARAQAHLAVLSGALSDQDLAIRPSGVTRRIVRAEA

Query Match 17.2%; Score 103.2; DB 1; Length 123580;
Best Local Similarity 53.5%; Pred. No. 0.074;
Matches 291; Conservative 0; Mismatches 238; Indels 15; Gaps 3;

Qy 72 GTGACCTGCGCGCTACCCGTTCCACGACGAACTTCTGGCTCCCGGTCCTCCCGGCG 131
Db 43432 GGGACACAGCTCGCGCGCTGCTGCGCCGACCTGCTCGCGGACCCGCGCGCAC 43491
Qy 132 CGGGTCCCGACACCGCGGACGAGTGGCTTACAGCTCGCTGCGACCCCGTCACTC 191
Db 43492 ACATCGGACACGACCGCTCGCGGACCGGAGAGTTGGAAACCGCTGTGCTGCGCGC 43551
Qy 192 GGGCGGCTCTCTCTGCGCGGACCGGCTCTGAGTGAACCGAGCG---GAGTACCCCGC 248
Db 43552 ACGCCGACACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43611
Qy 249 GCTTGAACGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 308
Db 43612 GTCCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43671
Qy 309 ACCGGGAGTGGCGCGCGCGGATCGCGCGGACCTGCGACCGCGCTGCGACCGCGCGCTG 368
Db 43672 GGGACCGACCGCGCGCGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43731
Qy 369 TCCACTGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43791
Db 43732 ACCGCGGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43791
Qy 420 CCCAGCTGACACCTTCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
Db 43792 GGCACACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43851
Qy 480 CTGTGCTGTGACCAAGGACGCG---CGCGCGCTGACCGCTGCGAGACGAGTGCATCCG 536
Db 43852 CTGTGCGCGCTGACCCCGCGCGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43911
Qy 537 GCCCAGGCGCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
Db 43912 GCCCAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43971
Qy 597 GGTG 600
Db 43972 GCGG 43975

RESULT 14
AX211739 125401 bp DNA linear PAT 06-SEP-2001
LOCUS Sequence 35 from Patent W00159126.
DEFINITION AX211739
ACCESSION AX211739.1 GI:15523950
KEYWORDS Streptomyces noursei
SOURCE Streptomyces noursei
ORGANISM Streptomyces noursei
REFERENCE 1 Zorchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaaset,T.,
AUTHORS Stroom,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
TITLE Gene cluster encoding a mysetacin polyketide synthase and its
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;

Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGE'S TEKNISKE HOESKOLE (NO) ;
ALPHARMA AS (NO) ; Solvent AS (NO) ; Zorchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaaset, Trygve (NO) ; Stroom, Arne Reidar (NO) ; Valla, Svein
(NO)

FEATURES
source Location/Qualifiers
1..125401
/organism="Streptomyces noursei"
/mol_type="unassigned DNA"
/db_xref="taxon:1971"
/note="ATCC 11455"

ORIGIN

Query Match 17.2%; Score 103.2; DB 6; Length 125401;
Best Local Similarity 53.5%; Pred. No. 0.073;
Matches 291; Conservative 0; Mismatches 238; Indels 15; Gaps 3;

Qy 72 GTGACCTGCGCGCTACCCGTTCCACGACGAACTTCTGCTCCCGGTCCTCCCGGCG 131
Db 43840 GGGACACAGCTCGCGCGCGCTGCTGCGCCGACCTGCTGCGGACCCGCGCGCAC 43899
Qy 132 CGGGTCCCGACACCGCGGACGAGTGGCTTACAGCTCGCTGCGACCCCGTCACTC 191
Db 43900 ACATCGGACACGACCGCTCGCGGACCGGAGAGTTGGAAACCGCTGTGCTGCGCGC 43959
Qy 192 GGGCGGCTCTCTCTGCGCGGACCGGCTCTGAGTGAACCGAGCG---GAGTACCCCGC 248
Db 43960 ACCCGGACACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44019
Qy 249 GCTTGAACGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 308
Db 44020 GTCCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44079
Qy 309 ACCGGGAGTGGCGCGCGCGGATCGCGCGGACCTGCGACCGCGCTGCGACCGCGCGCTG 368
Db 44080 GGCACACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44139
Qy 369 TCCACTGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44199
Db 44140 ACCGCGGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44199
Qy 420 CCCAGCTGACACCTTCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
Db 44200 GGCACACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44259
Qy 480 CTGTGCTGTGACCAAGGACGCG---CGCGCGCTGACCGCTGCGAGACGAGTGCATCCG 536
Db 44260 CTGTGCGCGCTGACCCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44319
Qy 537 GCCCAGGCGCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
Db 44320 GCCCAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44379
Qy 597 GGTG 600
Db 44380 GCGG 44383

RESULT 15
AX697991 11238 bp DNA linear PAT 02-APR-2003
LOCUS Sequence 15 from Patent W003010193.
DEFINITION AX697991
ACCESSION AX697991
VERSION AX697991.1 GI:29499041
KEYWORDS Micromonospora carbonacea
SOURCE Micromonospora carbonacea
ORGANISM Micromonospora carbonacea
REFERENCE 1 Farnet,C.M., Staffa,A. and Yang,X.
AUTHORS Genes and proteins for the biosynthesis of rosamycin
TITLE

JOURNAL Patent: WO 03010193-A 15 06-FEB-2003;
 FEATURES Ecopia Biosciences Inc. (CA)
 Location/Qualifiers
 source 1..11238
 /organism="Micromonospora carbonacea"
 /mol_type="unassigned DNA"
 /db_xref="taxon:47853"

ORIGIN

Query Match 16.7%; Score 100.2; DB 6; Length 11238;
 Best Local Similarity 52.9%; Pred. No. 0.33;
 Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1;

QY 143 CACCGGAGACGAGTGGTAAACAGCTGCGCTGGCAACCGGTGACCTCGGCGGTCTC 202
 DB 2949 CACTTGAAACCGTTGCCAGACGAGACCCCGACCTCAACACCGCTGTGTGAT 3008
 QY 203 CCTGGCGGACGGGTCCTGTGTGACCGGAGCGGCAATACCCCGGCTGAGCGAGT 262
 DB 3009 GGTGCCGAGGCGGTGCGGTGACGGGAGTGTGGCGGGGTACGAGAGCGCTGGCGC 3068
 QY 263 GGTCCGAGACGCGCTGGAACAGCGCGGGCGACCGTGTGTGACCGCGAGTCGCG 322
 DB 3069 GCGTGGGCGCTCCGTGCACTGACCGTCCGACCGCGACCGGCGCGGCGCTTGGCGC 3128
 QY 323 CGCCCGGATGGGCGCGCACTCGACCGCGTGAACGAGCACCGCCTGTCACTGTGTCTC 382
 DB 3129 GCTCTGACTTCGCGCGCGCCCGGACCGGAGCGGCGCGCGCGCGCGGTGTCTC 3188
 QY 383 TCTGCTCGCGCTGCGGAGCGGCGGTGTGACGACCCGACGCTG-----428
 DB 3189 CCGTGTGGCGCTGCGGAGCGGCGCGGACCGCGCGCACCGCGTACCGCTCGGCTGAC 3248
 QY 429 -GACACCTTCGCGTGTGTCAGGCGCTCGGCGGACCGGAGTGAACGTCCTGTGCT 487
 DB 3249 CGCTCGCTGCGCTGATCAAGGCAATTGCGGAGCGGAGGACGAGGCGCGCTTGGGC 3308
 QY 488 GGTGACCAAGGAGCGCGCGCGCGGTGACCGTTCGAGAGACGATCGGCGGCGGAT 547
 DB 3309 GGTACCGCGGAGCGGCGGTGTCCTCGGCGAGGTGCGGAGCGCGGCGGCGCA 3368
 QY 548 GGTGAGTGGGCTCGCGCGGAGTGGGCGTGAAGTCCCGCGCGGTGGGATG 600
 DB 3369 GGTGTGGGCGCTCGGCGGAGTCCGCGCTCGAATGCGGACCGATGGGCG 3421

Search completed: April 7, 2006, 16:57:34
 Job time : 2906.98 secs

This Page Blank (uspto)

gene
CDS

928..2061
/gene="megDVI"
928..2061
/gene="megDVI"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"
/protein_id="AAG1907.1"
/db_xref="GI:10179842"

translation="MAVGDRRRLRGELIQMARGLYMGFGANDLYSMLLSGRDDPWTTMYRLTAARGLVPASRAGTWVVDGDRRTAEVLADGFHTGPEDARMQWACPSASA
GPFRRFVARTSDASLVTDADMLOORCARLTVEIGSRPDLVNDPARREVPILAGTAA
LGKVPDPLRSWTSAITRYCDNAQYSPQQALVTBQALRLDESIDVTSGRRAATLVGV
ALELANVGMNVALATELPBELAARIADPEFATVTEVTSKSTSGVHLERTASDKR
VGVDVPPVGGVVVAANRDPEVFETDPDRFDVRGDGAELISRPDGSPTDIDLAV
ATLAAPARAAPVLPRLSRGSPVLRSSPVARGSRCPVEL"

2072..3382
/gene="megDI"
2072..3382
/gene="megDI"
/codon_start=1
/transl_table=11
/product="TDP-megosamine glycosyltransferase"
/protein_id="AAG13908.1"
/db_xref="GI:10179843"

translation="MRVVFSSMAVNSHLGLVPLASAFQAAGEHVVVASPALTDVTV
GAGLVAVGVGDVLEVMHAAGODIVEMRTLLMVWDOSHHTMSMDLIQMOTFTPT
PALMSPSLIDKMVPCSRMPDWIWMPLTPAAPLARVGTGHARMYLGPVAR
ARQSITRLIAHQEVRHEDPLAEMDTLRRFGDPHLSPBEELIGOMVTDPIPEPL
RIDCVSRTRIAQEVYHPNGSVVPMALRSEPRRRVCITIGSSRHGIQOVSIGEMLD
RIADDAEFVATFDQCLVGVGSPVANRTGTPVNAVLTLCATYTHHGTSQMULTA
AIHGVPQTIILSDPVEHAKOLQDLAGSLIPVGMTAEHLRGAILEVLDBPATRILCA
EMKRGMQIIISPAOVGICODLADRABRGORPRRTAPEHLPR"

3462..4634
/gene="megY"
3462..4634
/gene="megY"
/codon_start=1
/transl_table=11
/product="mycrose O-acetyltransferase"
/protein_id="AAG13909.1"
/db_xref="GI:10179844"

translation="MTSTINDTTARPALNSLTGMRFYAAFLVFPTHVLSRLINSTVY
VADGDAFMQTGVRGVSEFFTLSCFVLTWMSARASDSYWSFWRRVCKLPENHLVTFR
AAVFLVJAGVGEALIMILLIHAMPALKEISFGINSYMSLACAPFLYCFPLP
IFWISGIRBERLMAAVFAVIAVAVPADLLBSGPLRGLSEYALIDMWPLTYTPE
ATRSSLETIGILIARLITGRWINGLPAVILPVPFPVASLFLRGYALISSMMILP
LVLLIASGATADLOOKRTFMKRWVWMLGDVSPALYNHFLVITYGADLDLCSGTEDBA
PLGLALFMIIPELVALSVLSWLNYRFELPWRUNWARPASARRKATEPEQTPSRR"

4651..5775
/gene="megDII"
4651..5775
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-amino transferase"
/protein_id="AAG13910.1"
/db_xref="GI:10179845"

translation="MTTYWVSYLLEYERRADILDADVOKVPASGSLILGGSVENFETFR
VARHGIACVCGVNGTAVNKALSESVGGRDREVVTASNTAAPVLVILDEIGARPVH
VNVREDYLTMDLVEAAVTPPTTKAIVPVHLYGCQCVMWTLARELDNRGLKVDCQA
AGARDGDLGATGSDAAAFSFYPTKVIAYDGAGVVTNDDETALAIRLRURYGMEB
VYYTRTPCHNSRLDEVQELIRKLRTIDAVARRAVADORVYDGLADDSHGLEIE
PYVTGNHEVFPVVVYVRRHRDEELIKRLADGVDSLNTSYMPVHTTMGFALGVASG
SLPVTBLRAGEIFSIPMFSLPHDLQDRVYEAVERVITGL"

5822..6595
/gene="megDIII"
5822..6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"

gene
CDS
/protein_id="AA033912.1"
/db_xref="GI:10179847"
/translation="MPNSHSTSTSDVAPYEPADIVYDHFYHGKGVKAPRAADLVVEA
RKHPQATLILVACGTGSHLVEALDFKRYVGYDLSAAMLATRARDPRKELHOGM
RDFELDRFDVVCTGPFSSSTYLVDEALDFVADLANALACHLAPGGLVVEPWPFPETPR
GMVADIVTSGDRIISMSTHVPAGLFDPRASRTTHYTVGSSPAGIIEHTHEVMVNL
FAPAYAOAFORADLSISYGVHDLFSPLFVGVAAEPCR"
6592..7197
/gene="megDIV"
6592..7197
/gene="megDIV"
/codon_start=1
/transl_table=1
/product="UDP-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="AA033911.1"
/db_xref="GI:10179846"
/translation="MRVEBIGEVFTFTDQTFADRGVEGTAVQEDVFYAALGRPLF
PVAOSTTSGRRGVGHFTTTPGSKAKYVCARGAMPADIVRGSTPTFGABEV
ELASVWSTRLPLVPMGHLPFSLDDPTTLVYMGAGVVPKKEVPHLDDELALPIBA
DLDIVMSRDPVAPTLERADQGLIPYACRAAAHNVAT"
7220..8206
/gene="megDV"
7220..8206
/gene="megDV"
/codon_start=1
/transl_table=1
/product="UDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AA033913.1"
/db_xref="GI:10179848"
/translation="MVVLGASGFLGAVTTHALADLPYRVLVAREVVVPSGAVADYE
THRVLDLPRGLAEVADARAVFPFAOIRGTSMSRSESDVARETNVGLVRLIAY
LSRBPAAVVPFGSNTQGVTVTGAIVLDSGDHPHGVQDROKTEGOLIKKATAG
AIRATSLDLPVFGVPAPATGDDRGVSTMRALRTPQPLTHMDGTVRRRLIYYMDA
ARAFTALDHADLAGRHFLLTGRSMPLFPAVARSVAKHGTSDPVEVSVPPRA
HMDSDLSVSDVAPRFTAIVTGPRATVYMAEAVDRTVAAALAPRAAASPBS"
8220..9220
/gene="megDVI"
complement(8228..9220)
/gene="megDVI"
complement(8228..9220)
/gene="megDVI"
/codon_start=1
/transl_table=1
/product="UDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AA033914.1"
/db_xref="GI:10179849"
/translation="MGTTGAGSARVVRGSRALHTSRMLGTNFSRGVTTDDALRLMD
HALBEGVNCITDADIVGWRILYKGTIEELVGMFQGGRRRETVLATKVSEMSERVND
GGSAGRIAAVAAECSLRLLGVNDHIDLYOTHLIRAPAPMEVMDAAHLVSGSGVGVV
GSSNLAGNHIAAECSSARBNLLGMISHOCLYNLVAHRSPEVLAPOAVGVFAPAS
PLHGSLSGVLEKCACTAAVYVSAQGRADVLPAVRPLVAYEVDYCRLLGDAPEVGLA
WVLSFPLGLAVIGRPTREQDLSALRAHELTLGSEBELRELAITPAADVDPV"
9226..10479
complement(9226..10479)
/gene="megBV"
complement(9226..10479)
/gene="megBV"
/codon_start=1
/transl_table=1
/product="UDP-mycarose glycosyltransferase"
/protein_id="AA033915.1"
/db_xref="GI:10179850"
/translation="MRVLTSPFARHTPOGLVPLMALHTAGHDVAVASOPELTDVVVV
GAGLTSVPLGSDHRLFDISPPAAQVYHRYTTDLDPPARGPLRSMEPLGKLEKTSR
VFPVVNNDSFVDELVEFMDMRPDLVMEPFTFGAFAAAGAAHARILMGSBLTGY
FRSRQDRGRRPADRDPLDGLGMLTEVAGRFGADYSBLAVQMSVDQLEPESRLEP
GLESVHTTLLPNGSSVVPQWLRTSDGVRVCFTGYSALGITSNPOEFLTTATLAR
PGLFVIVRSGDLPSAVPDNRYLVDVPMNILLPGCAAVIHNGAGGSAWTAHHGVQC
ISVAHEMCTVARGORTAELGAGVLRPDRVADTLMQALATVVEDRSHAEAKLROE
ALAAPTPAAVVPFLEALAHQRADR"
10483..11424
complement(10483..11424)
/gene="megBIV"
complement(10483..11424)
/gene="megBIV"
/codon_start=1
/transl_table=1

Query Match	Beat Local Similarity	100.0%; Score 601; DB 1; Length 47981;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Query	1	GCCTCTACGCGGCGGCGCAACGCTTCTCTGACGCTTGCCTCCGTCTGTGCGCGCACT 60
Db	37506	GCTCTCTACGCGGCGGCGCAACGCTTCTCTGACGCTTGCCTCCGTCTGTGCGCGCACT 37506
Query	61	GGGCTGCCCGGTCTGACTGATCGCTGGGGGTCTGTGGGCGCGGCAAACTGCGCGTACC 120
Db	37566	GGGCTGCCCGGTCTGACTGATCGCTGGGGGTCTGTGGGCGCGGCAAACTGCGCGTACC 37566
Query	121	GAGGCGCGGCGCACTCTGCGGAGCGAGGCGCTGCGGCGCACTGACCCGCGAGCGGCGGATC 180
Db	37826	GAGGCGCGGCGCACTCTGCGGAGCGAGGCGCTGCGGCGCACTGACCCGCGAGCGGCGGATC 37826
Query	181	GAGGAGCTGCGGAGCAACCTTGAAGCGCGGGAACCGTGGGTGTGCGTGTGTGACCTTGAC 240
Db	37886	GAGGAGCTGCGGAGCAACCTTGAAGCGCGGGAACCGTGGGTGTGCGTGTGTGACCTTGAC 37886
Query	241	CGGGAAGCGGTTCTCTGAACTGTTCAACGCGCGCGCGCGCGCGCTTCTTGACGAATCTC 300
Db	37746	CGGGAAGCGGTTCTCTGAACTGTTCAACGCGCGCGCGCGCGCGCTTCTTGACGAATCTC 37746
Query	301	GGTGGGGTCCGCGCGCGGCGGCGGAGAGACCGGTCAAGAAATCGGATCTGCGCGGCGGCTG 360
Db	37806	GGTGGGGTCCGCGCGCGGCGGCGGAGAGACCGGTCAAGAAATCGGATCTGCGCGGCGGCTG 37806
Query	361	GCGTCTGATGCGGAGCGGAGCGCAACGTCAAGACATGTGCGCGCGCTGTGCGAGCGGAGGTG 420
Db	37866	GCGTCTGATGCGGAGCGGAGCGCAACGTCAAGACATGTGCGCGCGCTGTGCGAGCGGAGGTG 37866
Query	421	GCAGCGGTCTGTGGCGCAACGCGACGCGGACGCTGATGAGCGTGTGCGCTTCTGCTGAC 480
Db	37926	GCAGCGGTCTGTGGCGCAACGCGACGCGGACGCTGATGAGCGTGTGCGCTTCTGCTGAC 37926
Query	481	CTGGGATTTGCACTTCAGACCGCGCTGTGACCTGCGGAAACCGGCTCGCGCGGTGACCGG 540
Db	37986	CTGGGATTTGCACTTCAGACCGCGCTGTGACCTGCGGAAACCGGCTCGCGCGGTGACCGG 37986
Query	541	GTCGCGGTGCGCAACGACATGCTTGTGACCAACCGGACAGTGTGACGCGCTCAACCGGCGAC 600
Db	38046	GTCGCGGTGCGCAACGACATGCTTGTGACCAACCGGACAGTGTGACGCGCTCAACCGGCGAC 38046
Query	601	T 601
Db	38106	T 38106

product="TDP-4-keio-6-deoxyhexose 4-keioreductase"

/protein_id="PAGI3916.1"

/db_xref="GI:10179851"

/translation="MTRHVTLLIVSGFVSGALLREFTTHPLRLAVARTGSDQPGS
AGLEHLRVLDLLEPGRVAQVADTVVHLVAVAAGSTWRSATVEARVNAIGRD
LVALLRARPGRPAVLLPASTTOANPAAPSRVYOHKIBERLIROATEBGVDGYTL
LPALYHSGSPGGGTGRGVNTAMRLPALAGSLITMHGSGVRNLLAHEDVATAPFAL
HNHALLVGDVWTB8ADARPLGEIFETVAA5AVARQGNPAVPVSVPEPNANEDFR
SDDDSTSEFRTLTGMHPRVPLAEGIDRTVALISTYE"

12181..22821

/gene="megAI"

12181..22821

/gene="megAI"

/note="polyketide synthase"

/codon_start=1

RESULT 2

AX112026

LOCUS AX112026 47981 bp DNA linear PAT 01-MAY-2001

DEFINITION Sequence 1 from Patent W00127284.

AX112026

VERSION AX112026.1 GI:13938925

KEYWORDS

SOURCE Micromonospora megalomicea

ORGANISM
Micromonospora megalomica
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporinae; Micromonosporaceae; Micromonospora.

REFERENCE
1
McDaniel, R. and Volchegursky, Y.
Recombinant megalomycin biosynthetic genes and uses thereof
Patent: WO 0127284-A 1 19-APR-2001;
Kosan Biosciences, Inc. (US)
Location/Qualifiers
1. 47981

FEATURES
source
/organism="Micromonospora megalomica"
/mol_type="unassigned DNA"
/db_xref="taxon:47868"
complement (<1. 144)
/note="unnamed protein product; megVI (megT),
TDP-4-keto-6-deoxyglucose-2,3-dehydratase; SEQ ID NO: 2=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37805.1"
/db_xref="GI:13938926"
/translation="MGDRVVGHATPESTOSAIRFLTRHGPPATDDVHDMLHRAAR
HRE"

CDS
928. 2061
/note="unnamed protein product; megVII,
TDP-4-keto-6-deoxyglucose 3,4-isomerase,
TDP-4-keto-6-deoxyglucose 3,4-isomerase; SEQ ID NO: 3=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37806.1"
/db_xref="GI:13938927"
/translation="MAVGDRRRLRGRELQWARGLYWFGANGDLYSMLSGRDDPPWTW
YELRLAAGRPYASRAGTAVVGDRTAAEAETADPGFTGPPDARMWQVHCAASWA
GKREPRFATREDASATVADADQORCARLYTELSGRPDLVNDPAREPVHATGTA
LKVDDRLARSWTSAITRVCLDAQVSPQALAVTQALTALDEIDAATVGRPAATVAGV
AEIAANTVGNALVATLEBELARLADDEPATATRVTEKRTSPGVHLERTIASDRV
VGVDDVATGEEVTVVVAANRDPEVTPDRPVDVGDCAEILSSRSGSPRTDLALV
ATLATAPLRALAAVLEPRLSSGPIVRRRSFVARGVSRCPAL"

CDS
2072. 3382
/note="unnamed protein product; megI, rhodosamine1
transferase (eryCIII hemolysin), TDP-megosome
glycosyltransferase; SEQ ID NO: 4= translated amino acid
sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37807.1"
/db_xref="GI:13938928"
/translation="MRVVFSSMAVNSHLSFGLVPLASAFQAAGHREVVVSPALTDVT
GALITVAVGDDVDELVEFMSHAQODIVEWRTIDWDQSHTWSMDLLGMOTTPPT
FPLMSPPSLIDGMVFCRSMRPDIWMEPLTFAAIIARVTGTPARMLMGDVAR
AROSFRLIAHOEVEHRBDPLAEWPDWTLRFGDDHLSDEELVIGOMTVDDIPEPL
RIDTGRVTKMRYVPYNGSPVVPAMLRPERRRVCTLTGSGSRKENGICQVSGEMLD
ALINDIAEFVATFDDQLVGSGVSPANVTAGVPMNVLPTCAATVHNGTGSMLTA
AIGHQVQITLSDADTEVHAKQLODLGAGLSLVAGMTAEHLRGAIRIVYDEPAYRLGA
ERRRDEKRTDPSPAQVVICODLAADRAARGRPRTAAPHLEP"

CDS
3462. 4634
/note="unnamed protein product; megG (megY), mycarosyl
acyltransferase, mycarose O-acyltransferase; SEQ ID NO: 5=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37808.1"
/db_xref="GI:13938929"
/translation="WYTSRNLDTTAPRANSLTGKRFVAAPLVFPTHTYLSRLPNSVY
YADGLAFMOTGRGVSVFFLTSGFVLWTSASASVSVFMRKRVCKLPNNLVAF
AAVILVLTGOAVSGBALPNLLIHAPPALEISGINPVSSLSACAFAYICPLP
LAEVTSIRPERLMAAAVVAIWAIVVADLLPSSPLIPGLAESALQDWFLYTP
ATSELSFGITILARILITGRWNVGLPAVLLPVFPAASLPVAVLSSFWLTP
LVILISAGTADLQOKRTFMRNRVWVWLDVSGALPMVNRIVVVGADILSGSOTDA
PGLALFPIITPLAVSLVSMLLRVLELPMKRNARPARSARKKPYTBEOQPSRR"
/note="unnamed protein product; megDII, deoxysugar

CDS
4651. 5775

transaminase (eryCI, DnmI homolog),
TDP-3-keto-6-deoxyhexose 3-aminotransaminase; SEQ ID NO:
6= translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37809.1"
/db_xref="GI:13938930"
/translation="MTYVMSYLLERERADILDAVQKVFASGSLILQGVNPFETE
YARHGIAHCVGNDGNVAKLALSVGVGRDDEVTVSTAAATVLAIDIGARPVF
VVDVDEVDVMDTDLVRAATVPRKATVPAVLVGGCVMTALRELDGRGKIKVEDCAO
AHGARRDRLAGTMSDAAASFPTKYLGYCGGAAVTTDDTARLRLRYGHEE
VYVTVTGHNSRLDEVQASILRRKTLRLDAVYAGRAVAKRYVDGLADQDSHGLE
PVYVDGNEHVFYVYVVRPRDRIKRLRDGYDILSLIATYVWPHMTGTGAHGVASG
SLPYTERLAGEIFSLPMYPSLPDLQDRIVEAVEVITGL"
5822..6595
/note="unnamed protein product; megDIII,
daunosemethyl-N,N-dimethyltransferase (eryCVI homolog); SEQ
ID NO: 7= translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37810.1"
/db_xref="GI:13938931"
/translation="MENSHTSTSTVAPYPERADITYDFTYHGKGYRADAALVYA
RKTPQATLLDVACGTGSHLVELADSFREVGVGDAAMALATARNDPGRBLHQDM
RDFSIDRRFPVYTCMFSTGYLVDLBEALDAVANLAGHLVPGTLVVEPWFPEFEP
GMVGDILVTSGDRRISMSHTVPAGLPDPTASRMTIHYTVGSPRAGIEHTEVMTL
PARAYEAFORAGLSCSYVGHDLFSPGLFVGYAABPR"
6592..7197
/note="unnamed protein product; megDIV,
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryVII, dnmU
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID
NO: 8= translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37811.1"
/db_xref="GI:13938932"
/translation="MRYBELIGCVFTFTPTQFADERGVGTAYQEDVFAALGRPLF
PVAOVSTSRSGVGVGHFTPTMPGSMNAKYCARGRAMPAVDIRGSPGFGAEV
ELSAESVGLYLPVGMGHLFVSLDDTLVYAGVAPDERAVHPLDELALPIPA
DLIDMEBERVAPTLAEARDQGLIPYACRAAAHVVR"
7220..8206
/note="unnamed protein product; megDV, TDP-hexose
4-ketoreductase (eryBIV, dnmV homolog),
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37812.1"
/db_xref="GI:13938933"
/translation="MVVLGASGFGSANTHALADLPYRVLVAREVVVPSGAVADYE
THRYDLTERGLAEVADRAVFPFAOIRGTSGMRASEDDVARETNVGLVRLIIV
LSRSPHAPVYVPSGNTQVGRVTAAGVITDGSROHPRGVDRGHTGSOULKEKTAAG
AIRATSLRLPPVFPVPAAGTADRGVSTMRRLALTGQPLTFMHMDGTVRRLIYTTDA
ARAFAVTLADHADLAGRHFLGTGRSFWPLGSGVSRVSRHSGEDVPVSPVPPA
HMDPSDLSRVEDPARFPAVYGMRAVYTMAEADVRYAALAPRAAAPSBS"
complement(8228..9220)
/note="unnamed protein product; megBII-1(megDVII),
TDP-4-keto-L-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37813.1"
/db_xref="GI:13938934"
/translation="WGTGASARVVRGSAALHTSRMLGTVNSGRVTDODALRLMD
HALSRVACIDTADIVGMRLTKHTBELVGFCAQGRRESEVLAATKESMERV
DGLSARIVACENSRLRLGDHIDIYQTHIDRAAEREDVMAAELVSGSVGVY
GSSNLGWIHAAQESARMLLGMISHCYNLVAVRPELDVLPAAQAGVGFANS
PLHGILSGVLEKLAGTAVYAGCAQAVLLPAVRPLVEAEDVCRRLGAPAVGILA
MYLSRPGILGVITPRPEOLDSALRAAELTGESELELAIRPAVDDPV"
complement(9226..10479)
/note="unnamed protein product; megBV, mycarosyl
transferase, mycarose glycosyltransferase; SEQ ID NO: 11=
translated amino acid sequence"

/codon_start=1
/transl_table=1
/protein_id="CAC37814.1"
/db_xref="GI:13938935"
/translation="MRVILTSGFAHRTHQSLVPLAVALTHAGHDVRAVAGQPELTDVV
GAGLTVPSJSDRLRLDISPBAQVHRITTDIDPARRGRLSWETLHIBERTSRF
VFPPVNNDSFVDELVEPAMWRPDLVMEPFTAGVAAACAAHARLHMGSLTGY
FRSGODLNGORPADRPDLGQMLTEVARFGIDYSEDIAVQMSVDLPESRRLT
GLESVTRTLTPYNGSVVPQWLRTSDGVRVCTGYSALGITSNPOEPLRTATLAR
FDEIIVTSGPDLPAVSPDNVRLVDVPMNILLPGCAAVTHHGAGSMATALHHGVQ
ISVAHMDCLRCORTAEIAGGVFLRPDEVDATLTMQALATVVEDSHAEAKELRQE
ALAPTPAEVPLLELAHQHADR"
complement(10483..11424)
/note="unnamed protein product; megBIV, TDP-hexose
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;
SEQ ID NO: 12= translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37815.1"
/db_xref="GI:13938936"
/translation="MTRHVTLLGVSGFVGSALIREFTTFLRLRAVARTGSRDQPPGS
AGIEHLRVDLBEPRAVQVAVDIDVVHLVAAGSITMSAATVPAERVNAQIMD
LVNLAARBPAPVLAFASTOANPAPAPBRVYOHKTEARILLROATEGDVGVILR
LPAITGHSGSGCTGCGVVTAMLRPALAGEPITTMHSGSRRLIHYEDVATFTAL
HNHEALVGDVWTSGAUBARPLGIFETVAASVAKQGNPAVPVAVSPPEANAANDR
SDPDFSTEFVTLTGMPHRYPLAEGIDRTVALI,STYE"
12181..22821
/note="unnamed protein product; megAI; SEQ ID NO: 13=
translated amino acid sequence"
/codon_start=1
/transl_table=1
/protein_id="CAC37816.1"
/db_xref="GI:13938937"
/translation="MVDVPLDGTGRTPHPPLPMPPLCGHNEBELARAROLHAYLE
GISDDVAVGALAEETRAODPRAVAVASVTEALALAAORPSPSVYRGA
RPTAPVTVLPQGAQMPGATRLIASPVAAMRACTAPALAEVMSLTPYDGBR
HLRVEVQALTRAVQTSIALMRSFVRDADALGHSIGBLALVGAQDVEAAABA
AALMSRMEVLVGRDMAVALPAELAEAAVERDPAVAGVPSVLTGAPEP1
ARRVAELAAQVAAQVNVMAAHSAAVDAVABQMRSLTMFPAAGDDVPTATG
RLDTRBELGSGPFLPVPFDEATVAVLQGTGFTESPHVLAAISLQOTLDEVG
SPAIVPTLDRDQGLRPLLAVALAVTGVLTDMTAAVYGPVPGHLSAVAVTDDG
PSTFEDMAPDHYLRARLITVGAETALAGREVDAAATTRBLDVLVAQLATRLA
TAGGRDHLIMLDHPHALTEALNGPDESPRGSFTAEHDEPRAVAMAC
RLPGVYSPSEFMELEAGDAVGLPTDGDVERDLSFHPDPTSGRAHQDAQFLTG
ATSPDAFPGLSPEALAVRPOORTLELSEVLEVRAGIPTSLRTSGVFLIIPQ
EYGPRLAEGSGVGVLTGTTTSVAGRAVATLGGPAISVDTACSSSLVAVHLAC
QSLRSGSTMAALAGTVMPPTKMLVDFSRMNSILA PGRSKAFSAADGPRMARGAGM
LILRLSDARHGHFVLAIVRGTVANSBGANSCLSAVNGRAQVAVITQALASGLTPIH

Query Match 100.0%; Score 601; DB 6; Length 47981;
Best Local Similarity 100.0%; Pred. No. 6,5e-52;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTCTTACGCGCGGCGAACGCTTCTCTGACGCTTCCGCGTCTGTGTCGCGCGAGT 60
DB 37506 GCTCTTACGCGCGGCGAACGCTTCTCTGACGCTTCCGCGTCTGTGTCGCGCGAGT 37565
QY 61 GGGCTGCGCGTCACTGATGCTGCTGCGGCTGTGCGGCGGCGAGAACATGAGCGGTACC 120
DB 37566 GGGCTGCGCGTCACTGATGCTGCTGCGGCTGTGCGGCGGCGAGAACATGAGCGGTACC 37625
QY 121 GAGGCGCGGCACTACTGCGGACGAGGCGCTGCGGCGCATGACCCGCAAGCGGCGCATC 180
DB 37626 GAGGCGCGGCACTACTGCGGACGAGGCGCTGCGGCGCATGACCCGCAAGCGGCGCATC 37685
QY 181 GAGGAGCTGCGGACCACTTGGACGCGCGGGAACCCGTGGGTGTGCGTGTGTGACCTTGAC 240
DB 37686 GAGGAGCTGCGGACCACTTGGACGCGCGGGAACCCGTGGGTGTGCGTGTGTGACCTTGAC 37745
QY 241 CGGAGCGGCTTCGTGCAACTGTGTCACCGCGCGCGCGCGCGCCCTTGTGACGAATCTC 300
DB 37746 CGGAGCGGCTTCGTGCAACTGTGTCACCGCGCGCGCGCGCGCCCTTGTGACGAATCTC 37805

```
misc_feature
19..4470
/gene="eryA"
/note="putative"
/function="approximate span of module 3'"
97..1482
/gene="eryA"
/note="putative"
/function="approximate span of beta-ketoacyl ACP synthase
of module 3'"
1693..2670
/gene="eryA"
/note="putative"
/function="approximate span of acyltransferase domain of
module 3'"
3406..3921
/gene="eryA"
/note="putative"
/function="approximate span of beta-ketoreductase domain
of module 3, possibly non-functional'"
4171..4428
/gene="eryA"
/note="putative"
/function="approximate span of acyl carrier domain of
```

misc_feature module 3''
4471..10722
/gene="eryA"
/note="putative"
/function="approximate span of module 4''
4471..5847
/gene="eryA"
/note="putative"
/function="approximate span of beta-ketoacyl ACP synthase domain of module 4''
6034..7026
/gene="eryA"
/note="putative"
/function="approximate span of acyltransferase domain of module 4''
7165..9216
/gene="eryA"
/note="putative"
/function="approximate span of dehydrogenase and enoylreductase domains of module 4''
9433..9984
/gene="eryA"
/note="putative"
/function="approximate span of beta-ketoreductase domain of module 4''
10225..10483
/gene="eryA"
/note="putative"
/function="approximate span of acyl carrier domain of module 4''
10723..20235
/gene="eryA"
/note="ORF3 encoding modules 5 & 6 for 6-deoxyerythronolide B formation; putative"
/codon_start=1
/transl_table=1
/protein_id="AA26495.1"
/db_xref="gi:152695"
/translation="MSGDNGMTETKRLRYLKRVTYTLSDVYARLREVEHRAEPYATV
GMACRFPRDUDSPESFVSGGDALIAEADRGWEDPAPRIGLHAAAGDDAPG
FGISPREALAMPQORIMLETSMALRGPVLSRDEQAMQARADHILAREPRNSLRQGTTLAT
APDVLGVGTGTSVAGSVAAGRAVCLIGPAMTVDRACSGGLTALHLMESLRDRC
GLAAGVTVMSPGAFTEFRSGGLAAGRCRPFKXADFGIABGAGVYLQRLSA
ARRGRPVIALVRSAGVAVNOGASNGLTAPSGAORVIRALENAGVAGVVDVYEAH
GTRIDPILREVALISTYGABRDNDPITGSVKNSTGQAAAGVAGWKATIALR
HGMPTHTHPBPSPQIEMDGAIVSVSQAASWAPGERPRAGISRSISTNAHVY
EAPADBPBPAPDSGPVPLVSGRDEQAMQARADHILAREPRNSLRQGTTLAT
RRSAMERHAAVVGDRDIALAGRAVADGRIDRTATGQARTRGVAVPQAGWOG
MARDLRESGVFADSIDRCERALAAPHVMSLTDLSGAPLDRVVVOPALFAMVSL
AALWRSQVEPAVAVGSGRIIAAHVAGALTBDAAVLVRSVRLARLGGGQMAS
PGLTBOAERIGRPAGALISTAVNGPSVAGSGPDLDELIAECARIGITARRIPV
DYASHFOVESLREBLTTELGLISPVADVALYSTTQPIPTITMTQIYAYALREO
VRQDATRQLAEAGDAFVSPHVLTVGLIATLDSALPDADGACVGTLRKRGGL
ADFTALGEAAVAGVEVMSPAFADARPELVYPFORQYMLPIPTGRRARDBDDW
RYOVWREAAEWESALAGRVLLVTGPGVSELSDAIRGLGSGATVLTCDVESSTI
GTALBAADTDLSTVYLLSRDGAVIDPSLALVOLAAGAGAPLWVTRNAVVOY
ADGELVPAQAMWGICRVYGTIBQRPWCGIYDLVDAAGABIRSLAAYLADPREBOY
AIRADGIVARLVPPAPARATRMSPTGTVLTGSGTICAHVAMRLARSGASHVLL
GRGADAPGASBLREBLTALGTGTIACVADARLRLAVALAERABGRIVSAWHAA
GVSTPLDLDLEAFTEFIADVKVGTVMDELCPDLDAFLPSNAGWGSFGIASY
AAANAFLDGFAPRRRSGAPVTSIMGMAONMAGDGGAYLSOGILRAMPDPAVE
ELHTLHGOTSVSVMDRRRFEVLFRAAHRPLFDIAGARARSGRSEGPIAAOR
LAAISTARREHIALIRAEVAVLGHDDAIIIRDPAPRLGDSMTAYULRRLIA
VTGREKATVVFDEFTITRLADHYLRLVGALEAQAALVREPKADDTIAYVGA
CRPFGVANNPGLWFEFIVRGDAVTEMPTDGMIDLFPDPPQHHGTSYRHAFLD
GAADFDAAFGISPREALAMPQORIMLETSMALRGPVLSRDEQAMQARADHILAREPRNSLRQGTTLAT
OGYQODAVVPBDSBGYLLTGNSSAVSRAVVGIGJBPATVDRACSSSVLALHAC
GSLRGDCGLAVAGSVMAGPVETRSRGGGLAADRCARFASARFAPAGVAV
VILQRLSDARRAGROVLGVAGSAINOCGANGCLAAVSGVACQQRITRANRAGITGA
DVAAYEAGGTGRLDSPVBSALALATTIKRSGSSGPVLTGSKPIGHAQAAGVAGV
IKVVLGLNRGLVPPMLCRGERSPLIMSSGGVELEAASVPPADGVRRAGVAGV
SGTNAHVIIASBPBPPLPEBPQGVLAANSVPLIARSRTETLAAQARLIESAVD
SVPLTALASALATGRAHLPRRAALLAGDHEQLRQGLRAVAVAGVAPGATGTASAGV

VVFPQGAQMEGMAKGLLSVPVFASIBACDAVLSEVAGFSASEVLEQRPDAPSLER
VDVVOVLFEVMTSLARLMGACGVSPSAVIVAGVLSLEBQVRVALR
AKARLALAGGVNLSLAEPERARALIAPEBDSISVAVASPSSVSGDPBALAEV
ARCBDSVRAKTLIPVYASRSHVBEIRETIILADLOSARBAIPLYSTLHGRBG
ADMGPRYVYNLSQVRPDEARVAADGATVMSPHVETIADVAOITADVAIIS
LHRDTABEHLIAELARHVGVAVDNRNVPALPVALPYTPPEPQYHILAPESDL
ADSRVDMRPLATTVDLEGGFLVHSGESLSIAVERKAGRVVPASADRESAAL
REVPGEVAGVLSVHTGAATHALHOSIGEGVAPLMLVTSRAVVALGESBPVDEQAM

Query Match 66.2%; Score 397.8; DB 1; Length 20235;
Best Local Similarity 78.9%; Pred. No. 1.5e-31;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 1 GCCTCTACCGCGCGGCAACGCTTCTCGACGCTTCCGCGCTCGTGTGCGCGAGT 60
Db 14500 GCTTCTACCGCGCGGCAACGCTTCTCGACGCTTCCGCGCGCGCGAGGAG 14559
Qy 61 GGGCTGCGGTCACCTCGATCGCTGGGGTGTGTGGCCCGGACAAACATGCGGTAC 120
Db 14560 GCGCGCGCGGTGACGTCATCGCTTGGGGGCTTGGCCCGGACAAACATGCGCGGAGC 14619
Qy 121 GAGGCGCGGACCTACCTCGGCGGACGCGGCGCGCATGACCGCGGCGGCGGATC 180
Db 14620 GAGGCGCGGACCTACCTCGGCGGACGCGGCGCGCATGACCGCGGCGGCGGATC 14679
Qy 181 GAGGACGTGCGGACCACTCGACGCGGCGGACCGGTGAGTGTGAGTGTGACCTGAC 240
Db 14680 GAGGACGTGCGGACCACTCGACGCGGCGGACCGGTGAGTGTGAGTGTGACCTGAT 14739
Qy 241 CGGAGCGGCTTGTGTCGATCTGTTTCAACGCGCGCGCGCGCGGCTTGTGACGAACTC 300
Db 14740 CGGAGCGGCTTGTGTCGATCTGTTTCAACGCGCGCGCGCGGCTTGTGACGAACTC 14799
Qy 301 GGTGGGGTCCGCGCGCGGCGGCGGAGACCGGTGAGAAATCGAATCTCGCCGCGGCTG 360
Db 14800 GCGGTGCGCGGCGGCGGAGACCGGTGAGAAATCGAATCTCGCCGCGGCTG 14859
Qy 361 GCGTGTGTCGCGGAGCGGACGTCACGAGCATGTGCGCGGCTGTGTCGAGCGGAGTG 420
Db 14860 GCGGTGTCGTCGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGTG 14919
Qy 421 GCAGCGGTGTGCGGCGGACGCGGACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 480
Db 14920 GCGCGGTGTGCGGCGGACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 14979
Qy 481 CTGGGATTCGACCTCGACCGCGGTGACCTGCGGAGACCGGAGCGGAGCGGAGCGGAG 540
Db 14980 CTGGGATTCGACCTCGACCGCGGTGACCTGCGGAGACCGGAGCGGAGCGGAG 15039
Qy 541 GTCCGGGTGCGGACGACCATGTGTCGACCGACCGGACGAGTGAACCGGCTGACCGGAC 600
Db 15040 GTCCGGGTGCGGACGACCATGTGTCGACCGACCGGACGAGTGAACCGGCTGACCGGAC 15099
Qy 601 T 601
Db 15100 T 15100

RESULT 4
AR049368 AR049368 20235 bp DNA linear PAT 29-SEP-1999
LOCUS Sequence 3 from patent US 5824513.
DEFINITION
ACCESSION AR049368
VERSION AR049368.1 GI:6005407
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz,L., Donadio,S. and McAlpine,J.B.
TITLES Recombinant DNA method for producing erythromycin analogs
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;
FEATURES Location/Qualifiers

source 1. .20235
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 66.2%; Score 397.8; DB 6; Length 20235;

Best Local Similarity 78.9%; Pred. No. 1.5e-31;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1 GCCTCTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCTGCTGCTGCGCGAGT 60
DB 14500 GCCTCTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCTGCTGCTGCGCGAG 14559
QY 61 GGGCTCCGCTGCACTTCACTGATCGCTGAGGCTCTGTGGGCGGCGAACAATGCGCGTACC 120
DB 14560 GGGCGCCCGGTGACGTTCATCGCTGAGGGCTCTGGGCGGCGAACAATGCGCGGAG 14619
QY 121 GAGGCGCGGCACTACCTGCGAGCCGAGGCTTGCAGCATGGAACCGGCGAGCGGCGATC 180
DB 14620 GAGGCGCGGCACTACCTGCGAGCCGAGGCTTGCAGCATGGAACCGGCGAGCGGCGATC 14679
QY 181 GAGGCACTGCGGACCACTGCGAGCGCGGAGACCGCTGAGTGTGCTGTGAGCACTGAGAC 240
DB 14680 GAGGCACTGCGGACCACTGCGAGCGCGGAGACCGCTGAGTGTGCTGTGAGCACTGAGAT 14739
QY 241 CGGAGCGGTTGCTGCACTGTTCACTGCGCGCGCGCGCGCGCTTCTTCAAGCACTC 300
DB 14740 CGGAGCGGTTGCTGCACTGTTCACTGCGCGCGCGCGCGCGCTTCTTCAAGCACTC 14799
QY 301 GGTGGGGTCCGCGCGGCGCGGCGGAGAGACCGGTCAAGAAATCGATCTCGCGCGGCGCTG 360
DB 14800 GGTGGGGTCCGCGCGGCGCGGCGGAGAGACCGGTCAAGAAATCGATCTCGCGCGGCGCTG 14859
QY 361 GCGTCGATGCGCGAGCGCGCAACTCAAGCAATGTGCGCGCGTGTGCTGCGAGCGGAGTGTG 420
DB 14860 GCGCGCTGTGCACTGCGCGGAGAGCGCGGAGCACTGCGCGCACTGATCTCGCGCGGAGTGTG 14919
QY 421 GCAGCGGTGCTGAGCGCGCAAGCGCGGAGTGTGAGCGGTGAGCGCTTCCGCGAGC 480
DB 14920 GCAGCGGTGCTGAGCGCGCAAGCGCGGAGTGTGAGCGGTGAGCGCTTCCGCGAGC 14979
QY 481 CTGGGATTTGACTTCACTGACCGCGCTGTGACTTGCAGGAAACCGGCTGCGCGGAGCGAG 540
DB 14980 CTGGGATTTGACTTCACTGACCGCGCTGTGACTTGCAGGAAACCGGCTGCGCGGAGCGAG 15039
QY 541 GTCCGGGTGGCGACGACATCGTCTTTCAGACCACTGGAACCGGCTTCACTGCGGCGAC 600
DB 15040 GTCCGGGTGGCGACGACATCGTCTTTCAGACCACTGGAACCGGCTTCACTGCGGCGAC 15099
QY 601 T 601
DB 15100 T 15100

RESULT 5
AR095529
LOCUS AR095529 20235 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 3 from patent US 6004787.
ACCESSION AR095529
VERSION AR095529.1 GI:10023473
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.
JOURNAL Method of directing biosynthesis of specific polyketides
TITLE Patent: US 6004787-A 3 21-DEC-1999;
FEATURES Location/Qualifiers
1. .20235
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 66.2%; Score 397.8; DB 6; Length 20235;
Best Local Similarity 78.9%; Pred. No. 1.5e-31;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1 GCCTCTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCTGCTGCTGCGCGAGT 60
DB 14500 GCCTCTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCTGCTGCTGCGCGAG 14559
QY 61 GGGCTCCGCTGCACTTCACTGATCGCTGAGGCTCTGTGGGCGGCGAACAATGCGCGTACC 120
DB 14560 GGGCGCCCGGTGACGTTCATCGCTGAGGGCTCTGGGCGGCGAACAATGCGCGGAG 14619
QY 121 GAGGCGCGGCACTACCTGCGAGCCGAGGCTTGCAGCATGGAACCGGCGAGCGGCGATC 180
DB 14620 GAGGCGCGGCACTACCTGCGAGCCGAGGCTTGCAGCATGGAACCGGCGAGCGGCGATC 14679
QY 181 GAGGCACTGCGGACCACTGCGAGCGCGGAGACCGCTGAGTGTGCTGTGAGCACTGAGAC 240
DB 14680 GAGGCACTGCGGACCACTGCGAGCGCGGAGACCGCTGAGTGTGCTGTGAGCACTGAGAT 14739
QY 241 CGGAGCGGTTGCTGCACTGTTCACTGCGCGCGCGCGCGCGCTTCTTCAAGCACTC 300
DB 14740 CGGAGCGGTTGCTGCACTGTTCACTGCGCGCGCGCGCGCGCTTCTTCAAGCACTC 14799
QY 301 GGTGGGGTCCGCGCGGCGCGGCGGAGAGACCGGTCAAGAAATCGATCTCGCGCGGCGCTG 360
DB 14800 GGTGGGGTCCGCGCGGCGCGGCGGAGAGACCGGTCAAGAAATCGATCTCGCGCGGCGCTG 14859
QY 361 GCGTCGATGCGCGAGCGCGCAACTCAAGCAATGTGCGCGCGTGTGCTGCGAGCGGAGTGTG 420
DB 14860 GCGCGCTGTGCACTGCGCGGAGAGCGCGGAGCACTGCGCGCACTGATCTCGCGCGGAGTGTG 14919
QY 421 GCAGCGGTGCTGAGCGCGCAAGCGCGGAGTGTGAGCGGTGAGCGCTTCCGCGAGC 480
DB 14920 GCAGCGGTGCTGAGCGCGCAAGCGCGGAGTGTGAGCGGTGAGCGCTTCCGCGAGC 14979
QY 481 CTGGGATTTGACTTCACTGACCGCGCTGTGACTTGCAGGAAACCGGCTGCGCGGAGCGAG 540
DB 14980 CTGGGATTTGACTTCACTGACCGCGCTGTGACTTGCAGGAAACCGGCTGCGCGGAGCGAG 15039
QY 541 GTCCGGGTGGCGACGACATCGTCTTTCAGACCACTGGAACCGGCTTCACTGCGGCGAC 600
DB 15040 GTCCGGGTGGCGACGACATCGTCTTTCAGACCACTGGAACCGGCTTCACTGCGGCGAC 15099
QY 601 T 601
DB 15100 T 15100

RESULT 6
AY661566
LOCUS AY661566 32299 bp DNA linear BCT 05-NOV-2004
DEFINITION Saccharopolyspora erythraea polyketide synthase gene cluster,
complete sequence.
ACCESSION AY661566
VERSION AY661566.1 GI:55419414
KEYWORDS
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
REFERENCE 1 (bases 1 to 32299)
AUTHORS Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M. and
Santl, D.V.
TITLE Total synthesis of long DNA sequences: Synthesis of a contiguous
32-kb polyketide synthase gene cluster
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (44), 15573-15578 (2004)
AUTHORS Hu, Z. and Reid, R.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2004) Korean BioSciences, 3832 Bay Center Place,
Hayward, CA 94545, USA

FEATURES
source

gene
CDS

```
Location/Qualifiers
1. .32299
   /organism="Saccharopolyspora erythraea"
   /mol_type="genomic DNA"
   /db_xref="taxon:1836"
1. .10638
   /gene="eryA1"
1. .10638
   /gene="eryA1"
   /note="DBS1"
   /codon_start=1
   /transl_table=11
   /product="EryA1"
   /protein_id="AAV51820.1"
   /db_xref="GI:55419415"
   /translation="MADLSLSDBRTNQPRIVRPMPLSGCNESALRAROLRAHID
RPDAGVEVGAALAHDEQADAGPHRAVVAASSTSELDGLAAVADRPHASVVRGTA
RPSAPVVFVPGQQAQWAGAGELIGSRVFAAAMDACARAFEVTDWTLAQVLDSE
QSRVREVPQALFVQTSIALMRSFGVTPGA VVGHSIGELAAHVGGAADAAARA
AALMSRMIPLVNGDMAVALSADIEPRIDRDDVVLGA VNGPRSVLTSSEPV
ARVQELISAGVRAQVINVMGAASQVDIAGCMRSALAMPAGSGSEVPYASLTCG
AVDTRELVADYMRSRFLPVRPDAITSLAEVGTVPYASPHVLAALAQTLDAEG
SSAENVPLQRQGGMRFLAAQAFTGVAVDTVAIDVGAPEGLSEPLSEPARSEE
DEPAESGVNNAAPHVLERILAVNGETALAGREDAETFEELGLDVLAAQLA
KVSALGREVNIALLYDPTPRALAEALAGTEVAQRETRAKTEAAGSEPVAVANA
CLPGVSTPEEFWEILSEGRDAVAGLPTDRQMLLSMTWELRAGIIPRTSLASPTGVFGLIP
EATRPDAPFGMSPREALAVDPOORMLLSMTWELRAGIIPRTSLASPTGVFGLIP
OEYGPRLABEGSBVEGLMTGTTTSSVAGRIATLGLRGLISDTQLCSSLVANHLA
COSLIRGESSILAMAGSVTMEPTGMVDPSFMSNLADPGKAFSAAGNGMGAGG
MLLTERSDARNGHPVLA VLGTA VNSDASNGLSAPNGAOVRVIQOALAEGLAP
ADIDVAHGTTGLDPIEARLFEAYGDRQEPHLGSGVSNLGHTOAAAGVAVI
KMVLAVAGTGLPTLTHASERKEIDMSGAILSLDBEPMPAGAPRPAAGSVGIGI
TMAHILBEAPOVYGRVAGGVAPMVLSASABGLRQAARLAHLAHLRHPQDDR
DIAYSLTGRALPHRAAFAPVDSALRVLIDGLATNGADA VGTBAAQRAVFPV
GQWQWAGMADLIDTSPVFAALRECADLEPHLDSEVLPFLBAAREKODALST
BRVDVQPVMAVAVVNSLMSMRAGVPAATVGHISQEIATA CVAGALSDDARVVA
LRSVLIATMPGNKMASIIAAPAGEVRARIGDRVEIAA VNGPRSVVAGDSDELDRIVA
SCCTECIRAKRLAVDVASSHSHVETIDLAELAEIGEDPHLPGVFPFSVTYGTQTP
DELDAGWYRNLRTVRPADVRVLAACQYRTPLEVAHPILTAIIEIIGGSGADLS
AIIHLRGDGLADFGELSRAPAGVAVDMSVHLTGARVPLPTTPROREVMILS
PKPARSTEVDEVSALKRYLIMRPTAGSEPARLDGWLVAKTVAKTADTSTAREBL
BSGARVLELVVDARCGDELEARLSVGEVAGVLSLADVEAPEREREPFLASBLAT
LSLVQAMVASIELGCPIMTVTSVAATGFEFRVRAAGHLMGVGRVIALENPAMGGL
VDVPAGSVAEILARHLAAVVGSGAGEDOLALRDAVYGRWRVRAAPATDDMKCTGY
LVNCTGTCVGGQILRMLARBGAPHLILVSRSGPDADGAGVLAELRALKATVVAAD
VTDRSVRELLGGIGDVPVLSAVTHAAATLDDGVYDTITGERIRASARKLGRNHL
ELTRELDITAFVLFSFSAFAGAGLGGYAPGNAVILDLAQORS DGLPATVAMGTH
AGSMAEGPVADRPFRHGVIMPEPTACRALONALDRAEVCP IVIDRMDFLLA YTA
ORPTFLPEIDARRAAFOAAAPRVCALSLPAPERKALFELVRSRLAHLGASA
BRVPADQAFELGDSIALELRRLGAATGRLPTTVFDPHPVTRTLAHLAELCG
ATGABOAPATTTAPVDREPIATVGMACTLGRGVDSPERIMELITSGRSAEVPDRG
VPDELMASDAAGTRARHGNPAAGDPDAAPFGISPRALAMDQOQOALETTYALE
SAGIIPETTLRSDTGVFGMSHQGATGRPEREDGVGYLITNGTASVASGRIALVLS
LEGPALTVDTDCSSLVALHTACSLRNGDCGLVAGGVSMAGPEVETESRQGLS
PDGCKCPSPDEADFGIGEGSAFVLQCLSDARRGRGLRVLVAGSVANODGANGS
APSGVAAOORVIRBAMARAGITGADVAVVEAGTRGRLDPEVASALATLYKSGSSG
PVLGSKYSNIGHAQAAGVAGVTKVILGLERGVPMPLCGERSGLIDMSGGIELA
DGVKMSPALDGVTRAGCISAGVSGTNAHVITIAEPPEPFPQRRHLPATGVVPVL
SARTGAALRAQAGRLADHLAHPGLAPADVSTAKARQHREERA VLAATDRAVYER
LRVADGVAVVGVVTVGASLDGSEVFPVPGQAOWEGMARBLPVPAFBSIAEDCVL
SEVAGFSVEVLEPPADPESLSEVDVVPVFAVWVSLARLMRCAAPSAVSIACDGL
LEAAVAVGALSLEDKMGVVARAVAGSGMSLVGRSGRVEGLLDDMSGTG
LEVAAVVPDAVVVAGDQAAREPLFYCBGVCIGARBAIPUDVASTHAEVVRDETVQ
ALAGITPRRAEVPFFSTLGTGDLDETELDAGVTRNHLHVEFHSAYDALTDGQATF
IESPHPLVASSVDTDDASDAVLTGLRBDADADRFLTALDADRVAADWMEA
VLGAGLVLDLPGYFPGQRFMLPDRITPRLDGLMFPYRVTEVPRESEPARLKRML
VVVPGHEDEGTVTVRSGALAEAGPEVTVGGVLDGCAGVSLILAEQDAQVTL
VIVRGLDARGIDAPLWTVYRGA VVAGSEPVARPDQAKMGIGQVSLIRKPGWTGIVDL
PHMDPBLRGLTAVLASSBOVAVRALVARRRLSPHVTHTSTSYAVPGCTIIVTGC
TAGLGAEVARMLDAGRAHLALVSRRGPTGSGVDTABLTRLGARVVAHCDVSSRS
FVRBLVHGLIBEGDVAVGVVAAAGLPOOVAINDMDEAFDEVAAKAGAAVLDLBELCS
DAELFLFSSGSGVMSARQCYAAGNAFLDAFAHRRGGRGIPATSVAMGMAMAGMPT
GDBEAVSFLRSGVAMVPVPRALALDRLVLA SGETAVVVTVDVMDPAFASVTAARPR
```

gene
CDS

```
LDRIVTTASBRAGEPETESLDRDLAGLPRARTEAELVLTVTSATVLDHDDPKAV
RATTPEKLGFDLSIAAVRRLNLLNATGRLPSTLFDVHNAGAVGFLDAELGTER
GEAPSLAGLDALAEALPEVPATREBELVQRLERMLAALRPVQAADASGTGANPSGD
DYGAGVDLRLRGLRELDGD"
12080. .22783
   /gene="eryA11"
   /gene="eryA11"
   /note="DBS2"
   /codon_start=1
   /transl_table=11
   /product="EryA11"
   /protein_id="AAV51821.1"
   /db_xref="GI:55419416"
   /translation="MTDSKVAEYLRRTATLDLRARORIRELSDPATVSMACRLPG
GAVTPQRLMEILRGGETLSGPTDGMQDLARLHPDPDRGTSYVDKGFLDDAAGF
DAFPQVSPREAAAMPQORLLLETSWEIYENNGDILPHSLRGTATGVLGAARGBE
DTAAEDVSEISTYTGAPAVASRISTYMKLEBPSSISVDTRACSSSLVATLALESIRK
GESMSVVGGA VMAIPGVFVDSRQALADGRSRAFGAGAGFGFBSGVTLVLER
LSRARRGHVLA VVSGSLNDQASNGLSAPGAPQRRVIRQALBSCGLEPDQDVAV
BAHQGTALDGP IEAVALDQYGRDDADPRLPGVSKNIGHTOAAAGVGLIKVTL
ALANGELPATLHVREPTPHVMSGGVALLAQNQPMRGRGRTAAVSAAGISGTHNA
VYBEAPERHRETTADGRPVLVVVSABETALLRQAQIATLREPDADLGVGIG
LATTARABERAAVASTREAPRGLREILAGATLADAVTGCTEVDGRVYVFLPFGQ
LSQWAGMAGLELSSSPVFAKIRACDESMAPMDKVS DVLKQAPAPGLDRVDVDP
VLRAVMSLAEILMRSYGVSEPA VVGSGEIIAAHVAGSLTLEDAALVVGSRRLMS
LGEQGMMAAVALGEAAVREBLRPMODELSAA VNGPRSVVVGSGFGLARFSEBCALB
GIRVROIDVUYASHSPQIRBRERILLETTDIAIPRARVTFHSTVEBSMDGTDLDR
TYRRLRETYRTRPADTRLABSGYDAFIVSPHPVYVAVREABEADGADVVVGS
LHNDQGLSFLMSMATAHVSQVDIRNDVAPGAAPALPTTPFORRYTLQPAAPAA
ASBELARVSWTIEKESNGLDGLMLVLP LISPETEMLCATINANGRLACSDV
TSASRTEMAQVAVOAGTGFVGLSLSSDSACRPGVPAGAVGLTLVQAAGDAVDA
PMCLTCGAVRTPADDDIARPAOTTAGFPAQVAGELPGMWGVVDLPESVDALALIL
LVAVIRGGRAEDHILAVRQRLHGRVRYRSLPQSGSRMTPHGTVLTGAAASVVGQ
LVKRLMDRGAERLVLAGACRQDDILLAVEBAGASAVVCAVCAALREALDREPTALV
HAGLITNFGSISEAPBEERETLAKTALLALDDEVGDAVEREYCSSYAGI WGA
GMAVYAGSISEAPBEERETLAKTALLALDDEVGDAVEREYCSSYAGI WGA
AMTWERVVLAAGVSVVAVDVPVSEGPAAETPTALFRELARGQAAEPDPSGT
GAPORLAGSPBEOENLELVANAAVYLGHSSALEINVRASFSELGIDSLNAML
RKULASTGIRPASLIVFDHPTVTLAQLQHLRARLVGDQAA VVGAADBSSEPIAY
GICGRPFQGISPEQLRVYLAEGNLTGTPARDMGMTGLYHPDNPNGTSYVDKGG
FLTDADPFGPFOITREKLANDPQORLMERTAMEVBERAGIDPDALRGTDTVPFG
MNOQSVQMLAGEARVDYOGI GNSVSVISGR IAYFMEWGPALVLA DCSLSVLGI
HLMOQLRREGCSIALGAGTVMSDPTFVDFSTORQALDRKAKASARADNGLS
GVVALVLEPSLRANGHQVLA VLGAVNGQASNGSLA VNGSQCRVIRQALAAAG
VPAADVYVBARHGTGELDPIRAGAL IATYGDORDEPLRSGVKTNIGHTOAAAGA
GVTKVTLAMRGMPLRSLHDELISPHIDMSGANETVREVPVARGRPRRAGVSSRG
VSGTNAVITVEABEABEABARTERGP LPEVLSGSEAVVAAQORALAEHLRDTPELGI
TDAAMTATLARRARPDVAA VLGDRACVCAELDLAAGRSADVAAVTSAVPRVLV
FPQGAQWGMARLDLSSSEVFSEMSGRCAELSPHTDMLKLDVVRDGGDPDHERVD
VLOQVLP SIVMSLAEILMRAGVTPAA VVGSGEIIAAHVAGSLTLEAAAVVLA RQ
VARELDQGGAMSVGASRDELETLAVMDRCAVVAANVANGPQTSVAGPTABLDSEFPA
ABAREMPPRIATVRYASHSBEVARIERLAAELQTTIATVNGSVLHSTVTGRVITSA
MDASYRNLRLRPVLEQARGLVQEGFTDFVEYSPVILMAVESTRHAGAAVTCV
PTLARREGSPHEFLRNLIRAHVHGGAADLVRVAGSGPAELPTYPFHHORWPPPHR
ADVLSALVRGAHEHLILAAVDPHGGVATGRLSTGBOPALHVVGGRLTVGVSVL
VDLALASQHAQVLELVELVQRLPVLVAGALALRMSGVAPDESGRTITVYAAEDVA
DLDADQSGOAVTGLTLAGVAGAPRDTQWOPEDRIVRLPBDHYGLALBOGYEYGSFO
ALRAARKDSDSVYAEVSIADDEBYATHPVLILDVAQTLSTGLSGEGCGGLPRAMT
VTLHASGATSVRVATPAGDAMMLRTDPGHLVAVTDSILVTSGBEKKQFQEPRRG
EGEHLIASRETFGRKRREBRGSGGVNVNLSLGELELDESDLLADGSEVWAGKTD
LHDEIDASRETFGRKRREBRGSGGVNVNLSLGELELDESDLLADGSEVWAGKTD
LHDEIDASRETFGRKRREBRGSGGVNVNLSLGELELDESDLLADGSEVWAGKTD
HMSRGRVYKELVTOPAPVDQGVTLTGCGTGLGRLLARLVTENGHRLILVSRG
ADAGSDBELRKEITDLCASATIAACOTPDRAALALALDGLPRPLGVVHAAGVLDGL
VTSIDEPAVEVLELAKDVAANLLELANTGLSFTFVLPSSAASVLAGGQGVYAAANE
SLNLAALRLRRRGLPATALMGNAQASSEMTSGIGDR IARGLAALPTERLALPDSA
LIRGGEVVFPLSINRSALIRAEFVEVYLKGVNRAKTLRAAGAEAAAGVNVDRLAGRS
```


gene	SDQVAGIABIYRSHAAVSGYGSADQLPERKAFKIDJFDSIAAVEILNRLGTATVRL PSTLVDPDHPPLAVAEHLRLPAASPAVDIGRLDELKALRALSEDDHVGORL ESLRRMNSRRAPSTSAISDASDDELSEMDQRFGGSEDL"
CDS	22784..32299 /gene="eryA111" 22784..32299 /gene="eryA111" /note="DBRS3" /codon_start=1 /transl_table=11 /product="EryA111" /protein_id="AAV51822.1" /db_xref="GI:55419417"
Query Match	66.2%; Score 397.8; DB 1; Length 32299; Best Local Similarity 78.9%; Pred. No. 1.3e-31; Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Qy	1 GCGCTCTAGCGCGCGGCGGAAAGCGCTTCTCGAAGCCTTCCGCTCGTGGTCCGCGCAAT 60
Db	26561 GCGCTCTAGCGCGCGGCGGAAAGCGCTTCTCGAAGCCTTCCGCTCGTGGTCCGCGCAAT 26620
Qy	61 GGGCTGCGCGGCTCACTCGATCGCTGAGGCTGTGGGCGCGGCGGAGAACATGCGCGGATCC 120
Db	26621 GGGCTGCGCGGCTCACTCGATCGCTGAGGCTGTGGGCGCGGCGGAGAACATGCGCGGATCC 26680
Qy	121 GAGGGCGGCGACTTACTCGGCGAGCGGCTTGGCGGCGGAGAACCGCGCGGCGATC 180
Db	26681 GAGGGCGGCGAGTACTCGGCGAGCGGCTTGGCGGCGGAGAACCGCGCGGCGATC 26740
Qy	181 GAGGAGCTGCGGAGCAACCTTGGAGCGCGGCGGAGCCGCTGGGTGTGGTGTGGACCTTGAC 240
Db	26741 GAGGAGCTGCGGAGCAACCTTGGAGCGCGGCGGAGCCGCTGGGTGTGGTGTGGACCTTGAC 26800
Qy	241 CGGAGCGGCTGCTCGAATGTTCAACGCGCGCGCGCGCGCGCTCTTTCAGCAACATC 300
Db	26801 CGGAGCGGCTGCTCGAATGTTCAACGCGCGCGCGCGCGCGCTCTTTCAGCAACATC 26860
Qy	301 GGTGGGGTCCGCGCGCGGCGGAGAGCCGCTCAAGAAATCGAATCTCGCCCGCGCGCTG 360
Db	26861 GGTGGGGTCCGCGCGCGGCGGAGAGCCGCTCAAGAAATCGAATCTCGCCCGCGCGCTG 26920
Qy	361 GGGTGAATGCCGAGGCGCGAAGCTCAAGAGCATGTGCGCGCGCTGCTCGAGCGAGGTG 420
Db	26921 GGGTGAATGCCGAGGCGCGAAGCTCAAGAGCATGTGCGCGCGCTGCTCGAGCGAGGTG 26980
Qy	421 GCAAGCGGCTGCTGGGCGCGAGCGGCGAGCGGATGAGAGTGAAGTTCGCTTCGCTGAC 480
Db	26981 GCGCGGCTGCTGGGCGCGAGCGGCGAGCGGATGAGAGTGAAGTTCGCTTCGCTGAC 27040
Qy	481 CTGGGATTCGACTCGATGACCGCGCTGCACTTGGCGGAAACCGGCTCGCGCGGCTGACCGGG 540
Db	27041 CTGGGATTCGACTCGATGACCGCGCTGCACTTGGCGGAAACCGGCTCGCGCGGCTGACCGGG 27100

Qy	541 GTCCGGGTGGCGACGACCAATCGTTCGATCCACCGGACAGGTGACCGGCTCACCGCGCAC 600
Db	27101 GTCCGGGAACCGCGACGGTGTCTTGACCAACCGGACATCAACCGGCTCGCGGACAC 27160
Qy	601 T 601
Db	27161 T 27161
RESULT 7	
SEERYA	
LOCUS	9769 bp DNA linear BCT 14-OCT-1991
DEFINITION	S. erythraea eryA gene for 6-deoxyerythronolide B synthase.
ACCESSION	X56107
VERSION	X56107.1 GI:46975
KEYWORDS	eryA gene; erythromycin.
SOURCE	Saccharopolyspora erythraea
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE	1 Cortes, J., Haydock, S. F., Roberts, G. A., Beville, D. J., and Leadlay, P. F.
AUTHORS	An unusually large multifunctional polypeptide in the
REFERENCE	erythromycin-producing polyketide synthase of Saccharopolyspora
TITLE	erythraea
JOURNAL	Nature 348 (6297), 176-178 (1990)
PUBMED	2234082
REFERENCE	2 (bases 1 to 9769)
AUTHORS	Leadlay, P. F.
TITLE	Direct Submision
JOURNAL	Submitted (24-SEP-1990) Leadlay P. F., Dept. of Biochemistry,
FEATURES	University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
source	Location/Qualifiers
gene	1..9769
CDS	/organism="Saccharopolyspora erythraea"
	/mol_type="genomic DNA"
	/strain="NRRL 2338 (sub-species)."
	/db_xref="taxon:1836"
	57..9593
	/gene="eryA"
	57..9593
	/gene="eryA"
	/codon_start=1
	/transl_table=11
	/product="6-deoxyerythronolide B synthase"
	/protein_id="CAA39583.1"
	/db_xref="GI:46976"
	/db_xref="GOA:003133"
	/db_xref="UniProt/Swiss-Prot:Q03133"
	/translation="MSGDNGMTBEKRLRLTKRTVTELDSTVRLREYHRAEPIAIV GNACRPPGVDVSPESFMEFVSGGDAIEAPADRGPDPAIDGMLAAGPFDGAF FGISPREALAMPDPOORIMLEISWEALERAGHPVLSGATGVTGVDYGPRED ADDEVLYGVGTCTASSVAGSVAGVACGLGEGPMTVDYACSGTLHLAMESLRBEC GLAAGVTVMSPGAFTEFRSOGGLAADGRCKPSKADGGLAEGAVTVLQRLSA ARRERPVILAVLSAONOGASNGELTAPSGAQQRVIRTALEMSNGVRADVDYVAH GTGRLGDPTEIVHALSTYGAERDDPDLMTGSAVSNIGHTAAAGVAVMAVAVIL HEMERTLTFDEPSPOIEMDLGAVSVSGARSPAGBRPRAGVSFGISGTAHVIL BEAPADBEPAIDGSPVPLVSGDEQMRQAOARLADHLAASGTRCATPVSRBP AAAPBOVVGVGDRALAGLAVADRLADRTAGOGALPVSPPRVMVFPFGAONOGM AADLIREGQVADRSIRDCERALAHPVMSITDLSGAPRLDVPDVOPALPVMVSLA ALMRSHGVEPAVAVSHOGELIAAHVAGLTLEDAAKIVAVSRVLRIGGCGMASF GLSTQOABERIGRFAGALSIASVNSPVSADVVALYSTTGQPIDTATMDVAVYANIREQ VAFODATROLAEGADAEVBSPPHVLTVGIRATDLSALPADAGCVVGTTLRRDRGL ADFHRLGAEAVVAGVSVSGARVAVGLAGVTLVTCGVSELSDAIRSELGCGATVLT RDVRSRTTGRLAEDTALSTGVAVVAPHBEAVDPSIDALALVOLGAGVGAEPVAV VTRNAVAVADBEIVDPAQAMVGGIGRVAVVIBQPERKSGELVDLVADAAVSLAIVL ADPRGEQVAVIRADGIVKARLVAPAPARATHPLEBLAGTVLVYTGIGIAHLIRMLRSG AEHLVTLGRGADAPGASBELBELTALGTGTVIIACDVADARLBAVILAAEAAAGRT VSAVMHAAGVSTSLDULTEAEFTFIEIDVVKRGTVNIDELCPDDAFVLFSSNAGVW GSPGLASVAAANAFIDGPAPARASRGAPVTSIAMLWMGQNNAGDEGGEYLSQGLRA

mat_peptide

/gene="eryA"
/product="6-deoxyerythronolide B synthase"

ORIGIN

Query Match 63.0%; Score 378.4; DB 1; Length 9769;

Beet Local Similarity 77.8%; Pred. No. 1.6e-29;

Matches 470; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

```
QY 1 GCTCTCTACGCGGCGGCAAGCCCTCTCTGACGCGCTTGGCCGCTGCTGCGGCGAGT 60
Db 3658 GCTCTCTACGCGGCGGCGGCAAGCCCTCTCTGACGCGCTTGGCCGCGGCGGCGGCGAG 3917
QY 61 GGGCTGCGGCTGACCTGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 3918 GGGCTGCGGCTGACCTGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 3977
QY 121 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 3978 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4037
QY 181 GAGGAGCTGCGGAGCAACCTGGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 4038 GAGGAGCTGCGGAGCAACCTGGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4097
QY 241 CGGAGCGGCTTCTGGAAGCTGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 300
Db 4098 CGGAGCGGCTTCTGGAAGCTGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 4157
QY 301 GGTGGGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 4158 GGTGGGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4217
QY 361 GCGTGAATGCGGAGCGGCGGAGCTGCA---CGAGCATGTCGCGCGGCTGCTGCAAGCG 417
Db 4218 GCGGCGGCTGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4277
QY 418 GTGGGAGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
Db 4278 GTGGGAGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4337
```

QY	478	GACCTGGGATTGCACTTCATGACCGCGCTGACCTGGGAAACCGGCTGCGGCGGTGACC	537
Db	4338	GACCTGGGCTTGCATCTCATGACCGCGCTGACCTGGGAAACCGGCTGCGGCGGTGACC	4397
QY	538	GGGGTCCGGGTGGCGACGACCATGCTTTGACACCCCGGACATGACCGGCTTCAACGGG	597
Db	4398	GGGGTCCGGGTGGCGACGACCATGCTTTGACACCCCGGACATGACCGGCTTCAACGGG	4457
QY	598	CACCT 601	
Db	4458	CACCT 4461	

RESULT 8
SERRYABS
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
JOURNAL
COMMENT
FEATURES
SOURCE

1
Beyitt,D.J., Cortes,J., Haydock,S.F. and Leadlay,P.F.
6-deoxyerythronolide B synthase: cloning of the structural
gene, sequence analysis and inferred domain structure of the
multifunctional enzyme
Eur. J. Biochem.
2 (bases 1 to 20444)
Beyitt,D.J.
Direct Submission
Submitted (19-SEP-1991) D.J. Beyitt, Biochemistry Dept, Univ of
Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK
For related sequences see X56107 & M63677.
Location/Qualifiers
1..20444
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/strain="NRRL 2338"
/db_xref="taxon:1836"
/clone="SB55, DE10, PBK25"
/clone_lib="genomic in lambda EMBL3"
37..42
46..20268
/gene="eryA"
46..10749
/note="ORF B"
/codon_start=1
/transl_table=1
/product="6-deoxyerythronolide B synthase II"
/protein_id="CA44448.1"
/db_xref="GI:581651"
/db_xref="GO:003132"
/db_xref="UniProt/Swiss-Prot:Q03132"
/translation="MTDSKVAEYLRAATLDRARQRIRESLSDPLAIVSMCRLPQ
GNTFQRLMEILREGSTLGGFPDRGMDLRLHPPDPNGTGYVKGFLDAAAG
DAEFGPRLAAMDPOORLLRTSMELVENAGIDPHSLRTAGTAVFLGAKGYGS
DTAAEVEGYSVTGVA PAVASGRISTYMGEGSISVDTACSLSLVLAIVSLRK
GSSMAIVGAAMVATGVPFDRSQALADGSKAPGADGPGSEGVTLVLR
LSARRNGHEVLAVRGSALQDASNGSLAPSGPAGQRTVROLBSGLPGVDAY
EAHGTGALGDPIDANALLDYGRDRAADPLMGASVSNIGHTQAAGVGLNVVL
ALRNGELPATLVEEPTPHVWSSGVALAGNOPWRGERTRAAASAFOISGNH
VIVEAPREHERETTAGDGPVPLVSAASRAALRAQAQIAELERDADLAGVGLG
LATRAHERHAAVAIVASTREAVGLREIAGAATVAVGTVREVDRAVFLPFGQ
GSQAMGAKELISSPYFAGKIRACDSMAWQWKYSDVLRQARGAAGLDRVNVOP
VLFVWVSLAEIKMSYGVPAVVGSHGELIAAHVAGALTUEDNAKIYVGRSLRMS
LSGCGMAAVALGSAAVIERLRIPWODRLSVAIVNGPRVSVSGEPGLARAFSDCAAE

GINVRDIDVDYASHSPQIERREBELLETTGDIAPRPARTFHSITVSRSMDGELDAR
YWRNRLRETVRPADAVTRLAESGYDAFIEVSPHPVVQAAVEEADEADGEDAVVGS
LHRDGDLSAFLSMATAHVSVDIEMDVLPAPAPALPTYFORRYMLOAPAAVA
ASDELAERVSWTPRIKRESGNLDGDMVVTPLISPEKTEMLCEAIINANGRALRCYD
TSASRTEMAQAVNAQAGTGRGVSTLSSDSASACRPYPAAGVCLITTVOLGACATD
PVMCLTQAVRTPADDILAPQOTTHAGFQVAGLEPGRMGVVDLPESVDADALRL
LVAVLRGGGADEHLVARDGRHLGRVRVRSILFQSGRSWTPHGTIVLYTCAAPVGO
LVAVLRDGAERHLVLAGACPGDDILAAVEBAGSAVCAQDAALREALDEBVTALV
HAGTLNFGSISEVAPPEAFETIAAKTALAAVDEVYGRVAREVYCSSVAGIWMGA
GMAAAGSAYLIALAEHHRARGSCSWTPWMLPGAVDDGYLREBLRDLSDR
AMRTERVILAAGVSVAVDVDPVYSECPATRPILPHELAGGGQABESDSCGT
GEPQRIAGLSIPDEQENLLELVANAAEVLGHSAAELNVRASFELGDSITANAL
RKLSSATSTLPASLVFDPHTVALAOHRLVAGDQDAVRAVGAADSEPIAIY
GICRPGGIGSPEOLMRVLAEGANLTGTPARMGIDGLYHPDPNPCTSYVDGQ
FLTDADPDGFGGITPRBALAMPORMLETAMRABERAGIDPDALRDTDGVYV
MNGOSYMLLAGSABVDYOGIGNSASVLSGRITATTCBEGALTYDTRCSSLVGI
HLMOLKREBESLALAGVTVMSDPTTFSTQKSLADGCKASADPALSSE
GVALVLEPLSRARANGHOVLAVLRSAVNOQDASNGLAAPNGSERVIROLAAAG
VPADYDVAEHGTGTBLGDIKAGALITAYGODRDRPLGSKVTNIHTQOAAAGA
GVIXVILAMRHGMLPRSLHABELSPHIDMESGAVLEBEVMPAGRPBAGVSFQ
VSGTNAHVILEEAPABOBAVITGDRBAGVCABDLABGRSADAVAPVTSAPKPVY
TDAWTLATGRAPVYRAAVITGDRBAGVCABDLABGRSADAVAPVTSAPKPVY
FPGQAGWGMABDLLESSSEVFAESMSRCAULSPHTDWLADVAGDGDPEHEVD
VLQPLFISIMVSLAEIRAGAVTPAAVVGHSQGLAAHVAAGLSLEAAKVVALLSQ
VLEBLDQCGMVSAGSRDELFTVLARMQGRVAAVANGPSTVAGPTALDEFPAB
AEAREMKPRILAVRASHSEPVARIEDRLAEGLITAVAGSVPLHSTAYGEYDTSV
MEASWYRNLRPRVLPFGAVRGVIBGOFDFVVEPSPHVLIAVESTAEKQABEVTV
PTLRROSGPHERLRLNLAHNYGVADLAPVAGRPALPTYPRHQRFRPRPR
ADVSLGVRGAERHLLAADVPHGGAFTGLSTDBQWMLAEHVAGKTLVPSGLV
VDLALAGBDVGLPVLBELVORPLVLAGALLMSVGAAPDSGRITIDVAAEVA
DLAAMQSOHATITLAQVAGAPRDEOMPEDAVRIPLDDHYDGLAEQGYEYGFQ
ALAAWRKDSVVAEVSIADEBEGVAFHPVILDAVAVQTLISLAGPFGGKLEPPANT
VTLHAGSATSVRVATPAGDAMALAVTDPAGHVAITVDSIVVRSSTGKEQDEBPAG
EGEHLADWYRLABRGSTGRVAAADSDUAVRSGRPPEDAVIYVYBEGDPRRAA
RHSYMAALVYRMLEQEBELPGATLVATSGAVTSDDSVPEPAAAMVTRCAQA
ESDPFVLLDTPDEPMLPAVPDPQALTAGDVFPVRPLPSALTLTPAGTORLP
BGALIDSVAEPEADVEOPLRAGEVRVATVATGRVDVLLALGMYPKADMTZAG
VTVAVGPVDAPFAGSDVGLFQGAFAPIAVTBHRLARVPMQMSDADAVAAYIAATT
AHYALHDLAGIRAGOSVILHAAAGVGAVALABRAGAVALATAGPAGHGLRRLA
DDEHIASRETPARKFRRTGRCGVTVLNTSTGELIDBSADLLEDEGVFVEMKTD
LMDAGFRKRAPFDIGEDADDLSILBEVGLLGAEBLDRLEPVASMBELGSPALQ
HMSRGNHVKLVITOPAPVDPGTIVITGOTGLRLANHLVTEHGVHLLLVSRG
ADAPGSEDELAEBIEDLGAASAEIACOTRADLALSLDGLRPLTGVAHAGVLALQ
VTSIDPEAVOYLRAKVDAAMNHLTANTGLSFYLPSSASVLAGPQGVYAAANE
SLNALALERTGLPAKALKGMAGASEMTSGRIKARTGVALPTERBALAPESA
LRGGRVVPPLSTNRSALRAEFVEPVLKGMVAKLPAAGQAAGPANNVDRLAGSE
SDQVAGLAEVSRSHAAVSGYSGADLPKPKAKDGFDSLAAVELRNRLGATGVRL
PSTLVDPHTPLAVAEHLDRLPFAASPAVDIGDRLDELKALALSAEDGHQVQRL
BSLRRMNSRRADAPSTSAISEDASDDELFSMDQRFGGGIDL"
139.1416
misc_feature
/gene="eryA"
/product="beta-ketocacyl synthase"
1471.2763
misc_feature
/gene="eryA"
/product="acyltransferase"
4261.4455
misc_feature
/gene="eryA"
/product="acylcarrier protein"
4513.5787
misc_feature
/gene="eryA"
/product="beta-ketocacyl synthase"
5839.7116
misc_feature
/gene="eryA"
/product="acyltanferase"
7117.7563
misc_feature
/gene="eryA"
/product="dehydratase"
8743.9234
misc_feature
/gene="eryA"
/product="enoylreductase"
9418.10164
misc_feature
/product="ketoreductase"

misc_feature
CDS

10315.10509
/gene="eryA"
/product="acylcarrier protein"
10750.20268
/gene="eryA"
/note="ORF A"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase III"
/protein_id="CAA44449.1"
/db_xref="GI:46379"
/db_xref="GOA:Q03133"
/db_xref="UniProt/Swiss-Prot:Q03133"
/transl_table=11
GMACRPGGVDSPESTWERSGCGDAIARAPDRGHEPDDPDAALGCMALAAAGPDAG
FGISPREALAMPQOQIMLEISWEALERGHDEVSLSGSHYGVFTGVGVDPDE
ADEVLGVGTGTASSVAGRAVYCLGEGPMTVDTACSSGLTALHLMESLRBEC
GLIAGGVVMSPGAFTEPRSQGLIADGRCKPFSKADGFGIAGAGVLYQRLSA
ARREGPVLAVLAGSVNODGANSGLTAPSPQOQIRRLALENAGVAGDVYEAH
GCTGRLGDIIEVHALISTYGARDRDPDPLMTSVKSIHGTQAAAGVAGMKAVILAR
KSMPTTHLPDSRPOITENDLAVSVSQARSKVPRRPAKVSFSGISGTAHYIV
EAPPEADEPEPAPDSGPVPLVLSGRDEQMRQAGRLADHLAEPNNSLDTGFTLAT
RASAMEHRAVVVGDREDLALAGRAVADRIRADRTAGQGNRPRVAVMPFGQAQMO
GARDLIRESQVYFADSIIDCERLAPHVMSITDLISGARPLDRVVVOPALVAVVS
LALMRSHGVREPAVVGHSQGLIAAHVAGALTLEDAALVAVRSVRLRIGQGGA
SGIGTEQAQVSIIGRAPALSTASVNGPSSVILESGSPLELTAECABAHKARI
PDYASHSQVBSIREBELTELALGISPVASDVALYSTTGQPIDTITMTATYANIR
EYRFODATROULEAGFDPAVEVSPHVLVIGLEATLDLADPAACAVGTILRNRG
GLADFTLALGEVVAQCEVDMWSPADAPVLEPVYFQRVYMLPTFGGARBEDDM
RQVWRERAEWESASLAGVRLVVTGGSVEFSEDAIRSGILEGATVLTCDVRESRT
GVALLEADDTALSTVGAVPHEAVPDLALVALQALAGVBAVLTWTNRNAVY
ANGELVDPADNAGVIGRIVVGIIBOPRMGGLVLDVADAASTRSLAAVLTGREGVY
AIRAGCIKVALVPAAPARATPRLACTVLYTGTGCIHANLAWLRASGAEHLVL
LARGDAGASLEBELFALGTGVTIACDVADRRLAIVLAABASRTVSAVH
MAGVSTSELDLDTLEAFTEIADVKNRGLVNDLCPDIDAFVLFSNAGVMSPGIA
VYAANAFLDGPBARABEGAVTSLAMGLGONNAGGEGGYTLASOGIRAMPRA
VEHLHTLDGQTSVSVNDPDRRFPVELFTLAHRRLPBEIAGARABARQSEGPALA
QTLALICGRERHHLIRAEVAVYLGHDAAIDRRAAPDGLPDMSTVYDLNR
ILAVTGVREAAVVPDHPITTLADHYLERLYGAABEAPALVREVPFPDADPILAI
VMAKCPFGGVNPELMEFIVGGGDAVTEMPLDRKMDIDLFPDPORRGTSYSRHG
ALIDGAGDDAFLPGISPREALAMPQOQOQVETWELPEMNGIDHVSAGSPTGVL
GAAYQGGDAVVPEDSEGYLLTGNSSAVSQRVAVVLEBGAIVTVDRACSSVAL
HSAQVLRDCCGLVAGGVSVAGVPEVTEBSROGLAVDRCKAFSABDQFGLBE
GYAVUOLRPLSGPBARQVYGVVAGSAINDDGATNGLAASGVACQVIRAKARA
GITADVAVEAHNGTIRLGDVEBASALATTGSEBSGSSPVLSVKSNIQAAQAA
GAVYIKVVLGNRLVPEPMLCRGERSPLIEMSSQGVLEAVLSPMPADGVRAQAV
SAFVGSYTAHYTIAEPEPEPLPERGPGVGLAANSVVLISARTETLAAQARILE
SAVDSVPLTALASLATGRALPRRALLAGDHEQLRQLAVRAGVAPGATGTA
SAGGVVYVPPGQAGMEGARGLLSVVPASIRACDRLSVAGSASAEVLEORDA
PELBERVDVQPLFESVMSLALMKACVSPSAVIGHSGETIAAVAVATLSIEDCV
VVALRAKALRALAGKGVSLAPGERAALAPWEDRISVAVNSPSSVVSGDEBA
IAELVACDEBVARAKTLPVDAHSRHEBELRETLADLIDISARBAIPLYSTLHG
BRMDGPRWYDNLRSOVPDEAVSAQSPDGATFVMSPHVLTVADVEIADAAVAI
GSLHRDTAEHLIAELARAHVGVADVDMNVPAPVPLPVYFPEPOKYMALAEVSD
QLADSRVYDMRLATTPVDLEGGLVHSGASBSLTSATKAGVAPVVASARELAA
ALREVPGERVAGVLSHYTGAANALAHQSIGRAGVAPLTVTSRAVALRESPEDEQ
ANVWGLGRVWGLTEPERMGGLVDLPABRPBGGEVAPACLGADGHEDVVALDHAYG
RRLVAPRLGTRESWEPAGTACTVGGTAEOLSGVLAHRLARCEVBLVYVSRGVAIPA
AAELBAEIVALLGPKTITITACVADABEOLSKLEBELRGQCRPVRTVHTGVESRBLP
BIZGLIESVCAAKVYTGARLDELCPDAFTEFVLSGSGVWSANLGAVSANVYLDLTA
HRRRAREGRATSVAMGAMAGBMATGDEGLRRLRGIRPAPAPRAIRALHQLDNGCT
VSIADVWEAPAVGFTTAARPRPLDELVTTPAVGANVAYOAAAPAREMTSEBLEPETHS
VVAITGHSPPDAVGGDOPFTLELGYDLSVGRNQOQOTGTLAPVTLFEPHVARL
ADHIQOQDSTGTPARBASALNDGRQGVAGRSVYDLNLGLDPERHAPGSDPS
IDLVMAGQGEVTVYICAGTAIAGPHFTLALMALGIIVARVAPQGYEGBPLP
SSMAVVAVVOADAVIRTOGDKFPVAGSAGSANTALATLELDGHPRGVAVLIDVY
PBGQDANMANMLESLLATLPDRERTVMDTRTLTGAOVLRGVORPRETGSLTLLVS
AGRPMPDPSQKRTVPEHDTTAVVPDHPITVWGHADATARHIDAMLGNS"
10867.12111
/gene="eryA"

Query Match
Best Local Similarity

63.0%; Score 378.4; DB 1; Length 20444;
77.8%; Pred. No. 1.3e-29;

Matches 470; Conservative 0; Mismatches 131; Indels 3; Gaps 1;			
Qy	1	GCCTCTTACCGCGCGGCGCAACGCTTCTGAGAGCCCTTCGCTCGTGTGCGCGGAGT	60
Db	14533	GCTCTTACCGCGCGGCGCAACGCTTCTGAGAGCCCTTCGCGCGCGCGCGGAGGAG	14592
Qy	61	GCGCTGCGCGTCACTCGATCGACTGCGGCTGTGTGGCCCGGCGAGAACATGCGCGTACC	120
Db	14593	GGCGCGCGGTCAGTCATCGCTGCGGCGGCTGTGGCCCGGCGAGAACATGCGCGGAGAC	14652
Qy	121	GAGGCGCGCGACTTCTGCGGAGCGGCGCTTGGCGCGCATGACCCGCGAGCGCGCATC	180
Db	14653	GAGGCGCGCGAGTACTTGGCGAGCGGCGCTTGGCGCGCATGACCCGCGAGCGCGCATC	14712
Qy	181	GAGGAGCGCGGACACCGCTGAGCGCGGCGGACCGGTGGGTGTGATGATGACCTGAGAC	240
Db	14713	GAGGAACTGACATCACTCTGACCGAGTCAAGTCAGTCGTCGTGTGTGTGATGATGAT	14772
Qy	241	CGGAGCGGCTTCTGTCGAACTGTTCAAGCGCGCGCGCGCGCGCGCTTCTTGAAGAACTC	300
Db	14773	CGCAGCGGCTTCTGTCGAGCTGTTCAAGCGCGCGCGCGCGCGCGCTTCTTGAAGAACTC	14832
Qy	301	GCTGGGCTCGCGCGCGCGCGCGCGAGAGACCGGTCAGGAATGATCTCGCGCGCGCTG	360
Db	14833	GCCGCTGCGCGCGCGCGCGCGAGAGCGCGCGCGCGCGCGCTGCGCGCGCTG 14892	
Qy	361	GGCTGATGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGAG	417
Db	14893	GCGCGCTGCTGTGCGAGCGCGCGAGCGCGAGACCTTCTGCGCGCGCGCGCGAG	14952
Qy	418	GTGGCAGCGGCTGTGGGCGACGCGACCGCGCGCGGTCGATGACGTCGCTTCCGT	477
Db	14953	GTCCCGCGCGGTCGTCGCGACGCGGCGAGCGAGCGCGCGCGCGCGCGCTTCCGC	15012
Qy	478	GACCTGGGATTCGACTCCATGACCGCGCTGCACTTGGCGAAACCGCTCGCGCGTGAAC	537
Db	15013	GACCTCGGCTTCGACTCCATGACCGCGCTGCACTTGGCGAAACCGCGCTCGCGCGTGAAC	15072
Qy	538	GGGGTCGGGGTGGCGACGACCATGCTTTCGACCAACCGCGAGTGAACCGGCTCAACCGCG	597
Db	15073	GGGGTGGGGGAGCGCGACGAGTGTCTTCGACCAACCGCGACATCAACCGGCTCGCGAC	15132
Qy	598	CACT 601	
Db	15133	CACT 15136	
RESULT 9	31531 bp DNA linear SYN 03-NOV-2004		
AY771999	AY771999		
LOCUS	AY771999		
DEFINITION	Synthetic construct DBBS polyketide synthase gene cluster, complete		
ACCESSION	AY771999		
VERSION	AY771999.1 GI:54778584		
KEYWORDS	synthetic construct		
SOURCE	other sequences; artificial sequences.		
ORGANISM	1 (bases 1 to 31531)		
REFERENCE	Kodumal,S.L., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and		
AUTHORS	Santi,D.V.		
TITLE	Total synthesis of long DNA sequences: Synthesis of a contiguous		
JOURNAL	32-kb polyketide synthase gene cluster		
PUBMED	15496466		
REFERENCES	2 (bases 1 to 31531)		
AUTHORS	Reid,R.C., Kodumal,S.J., Patel,K.G., Menzella,H.G., Welch,M. and		
TITLE	Santi,D.V.		
JOURNAL	Direct Submission		
FEATURES	Submitted (05-OCT-2004) Kosan Biosciences, 3832 Bay Center Place,		
SOURCE	Hayward, CA 94545, USA		
	Location/Qualifiers		
	1..31531		

/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
/note="constructed for expression in Escherichia coli of
three subunits of the polyketide synthase from the
erythromycin biosynthetic cluster of Saccharopolyspora
erythraea"
21..38
/note="T7 promoter"
41..65
/note="lac operator"
72..10750
/gene="eryAI (morphed)"
72..77
/gene="eryAI (morphed)"
86..10750
/gene="eryAI (morphed)"
/note="similar to Saccharopolyspora erythraea EryAI of
GenBank Accession Number AY61566; contains His-tag"
/codon_start=1
/transl_table=11
/product="EryAI (morphed)"
/protein_id="AAJ39550.1"
/db_xref="GI:54778585"
/translaction="MADLSKLSDSRTAQPRIVRPPPLSGCNBSALRARAROLRAHLD
RPPDAGVGGALALDEODAPRAVVAASSTSLDGLAAVADGRPHASVVRGVA
RSPAPVFPVPGGAQMAQAGELGSRPVAAMACARAPRDMTLAOLVLDSE
OSRREVVOPALRAVGTSLAALRSRGTVPDVAVYSGISGLAAAHGCGAADAARA
AALMSRMTPLVANGDMAVALSADLEPRILAMDDVVLVANGSPSVLTSPREV
ARRVOELASGAVAAQVINTVMAASHAQVDIAGCMRSLAMFAPGSEVFPVSLTGG
AVDTRELIVADYMRSEFRLPVRFDAIETSLAEVPGTFVEASPHVLAALQOQLDAG
SSAAPTTLQROGCGMRFLAAQAQATVAADMTAAVDDGAPGSLPEPAPSE
DEPASGDMNAPPHVLERILAVINNETALAGRADATFRELGLDVLAAQLRA
KSGAALIREVINILIDHTPPRALLABLSGCTVAQRETAARTTEAAGRPVAVMA
CRUPGVSTPESFWEBSBGRDAVAGLPTRGMDLSEPHDPTRSGTAHQGGELT
EAPAFDFPFGMPREALAVDPQRLMLRLSMWELEKAGLPVSLQASPTVGGLIP
QEYGPRLABGEGVEGYLMTGTTVSAGRIAYTLGLEGPALIVDTACSSSLVAHILA
COSLRBESSLEAAGAVTVMPTQMLVDPRMNSLPDGRCKAFSAAGFMGAAGAG
MLLERLSARBNHGPLAVLRGTAVNSDASNGLSANPRAOVRVTOALASGLAP
ADIDVBAHGTGRLEDPTRARALFPAHYGRDREOPHLGSKNLTGHTQAAGVAVIT
KNVLAAGTLPPTLHASERSKSIDNSGASISLDEPEPPAGARPRRAYSSTGIG
TNAHALIERAPOVVERGERVEAGVAVAWVLSASABGLRQAANLAHLRHPQDER
DIAVSLATGRALLPHRAAFAPVDSBALRYLDGLATGNADGAVTSGRAOQRAVFP
GOGKMGAMAVDLDTPSPFAALRSCADALEPHLDEFEVIFPLAESARESDALST
ERYDVVQPMVAVWVSLASMMRAHGVBAVITGHSOGEIAACVAGALSIDDAVVA
LRSRVITMPGNKMASIAAPAGBYRARIQDRVETIAAVNPRSVVAGDSDELRLVA
SCTEGCTRRARLAVTASHSSHETIDALHABLGDFHPLPGVFPSTVTGATWP
DELDAGTWRNRLRTVAFADAVRALAAGVRTLEVSAPILTLAIEIRIDSGADIS
AHSLRGDSGLADFGALSGRAFAAGVAVMESVHLGTGARVPPLPYFORBERVMIQ
PKPARSTSTVEDVSALRYRIEMRPTGAGPARLDGTWLVAKYAGTDESTAREAL
BSAGARVRLVVARCGRDELARLRLRAGVAGVTLALLAVDABAREPVALIADLT
LSLVQANVSAELGCEPLMTTESAVATGTPFRVYRNAAALMGVRYALBNPVMGL
VDVPAGVABELARLAAVSGAGBEDLALRADGVYRRVNRVRAAPRTDBMRGTGV
LVTCGTGCGQOIALRWLARRGAPHLTVSBSGPDADAGELVALBELGARVTAAACD
LVDBESVRELIGIGDDVPLSAVFHAATLDDGDDVLTGERISERARAKVLRANLH
ELTRDELTLFVLESSFASAFGAPGLGVAGVAGNAYLGLAQOORSDGLPTAVAMGTW
AGSGMAGCPVADPRBRGCVIEMPERTGALONALDABAEVCIPTVDYMRFLAAYTA
QRPTRLDEIDDRARAPOLAAERFRVVALSLAPRPERKALFEIYRHHAAVLGHABA
ERVVADDAFRELGVDSLSALELRNRALAGVRLPTTVTDHPVRLAHLAALBSAS
ATGAEOAAPATTAPEVDPRIAIVGMACLPDSVSPRLMELITSGRBSAELVPDDBK
VPDLEMSDAAGTBAHGNFMAGADDAFAFGISPREALAMQOQALQOQALQOQALQO
SAGIPETTLGSDTVGVMSHOGVATGRPREPGVGYLLTGNTAVALSAGRIAYVIG
LBSGALVVDTRACSSLYALHTACGSLADGCGGALVAGVAGSAPVETTESRGCALS
PDGCRKPSDSADDFGLGBSAFVVLRLDSARBRGRVZGVAGSVANODGANGIS
APSGVAOQVYIRRAMARAGITGADVAAVEABGSTRIGDPEVASLATTYKSGSSG
PVLLGSYKNSIGHQAAAGVAGVATKVLIGERGVAPMLCNGESGILDSGSGIETLA
DGVEMSPADGVBRAGVSAFVGTNAHVIAAPRPEPVPQGRMLPATGVPPVVL
SARTGALRAQAGRLADHIAAHPGIATGAVSWTARAROHFEBAALTAADTAAYVR
LRVADDAVVPVYTGASDQGSIVFPFGCAQOMRGMARLELLPVYPAESIAEDDAVL
SEVAGFSVSEVLEPRDAPSLERVDVQVPLFAVMSILARLMRCAGVPAVITGHSQ
EIAAIVAGALSLDEGMVVARSRRAVRAVAGRSMLSVKRGSDVDELTLADDSWTGR
LEVAAVNGDPDAVVAGDAQAAREFLEYCEGVIGRARAI PVDYASHTAHVEPRDELVO

ALAAGTERRAEVFPSTLTGDFLDGTELDAGWYRNLRHFEHSAVOALTDGATATP
LEVPSPHVLASVOETLDDASDAAVLTGLEBDAGDRLFTLALADHTEGVNDMA
VLGRAGVLDLPGYPOGKRFMLQDRTTPRDELDMGFYRDLTMDPEVPSBEALRGRL
VVVPEGHEGCMTVVRSALABAGABRYTRGUGLVGDCAVSSILALGDCGAVOTL
VLYRELDABEGIDAPLMTVTTFGADVAGSVARPOAKIMGAGQVASLERGPMGLVDL
PHNPDELRRLTAVALAGSDQVAVRADARARLRSPAHYATSEVAVPGTLLVYTG
TAGIAGVARMILMARGAEHLALVSRGPDTEGVDLTAELTRLAGAVSAVCSVSE
PVRELVHGLIEQDVGAVVHAAGLPOQVALINMDDEAPDEVVAAKGAGVHDLCS
DAELFILSSGAGVWGSAQGAQVAAAGRAFLDAPARHRGRLPATVAMGLMAAGCT
GDEBAYSLERBCVRAMPVPRLAALDRVLASGTAHVYTVDMPAEYVLAAPRP
LIDRITVTASERAGEPETSLEDRDLGLPRAKTALVRLVTRSTAVYGHDPKAV
RATTPKEKELGDSLAAVRLNLTNAATGALPSTLVEHPNANASAVAGFDABLSER
GBAPSLAGLDALAEALPEVPATERBELVQRLRMLAALRPVAQADASTGANPSD
DLCEAGVDELLALGRELDDGNSHHHHH"
10815..10861
terminator
/note="T7 terminator"
10882..10899
/note="T7 promoter"
10902..10926
/note="lac operator"
10933..21677
/gene="eryAII (morphed)"
10933..10938
/gene="eryAII (morphed)"
10947..21677
/gene="eryAII (morphed)"
/gene="eryAII (morphed)"
/note="similar to Saccharopolyspora erythraea EryAII of
GenBank Accession Number M661566; contains His-tag"
/codon_start=1
/transl_table=11
/product="EryAII (morphed)"
/protein_id="AAV39551.1"
/db_xref="NC_54778586"
/translation="MTDSEKVAEYLRATLIDLPAAQRITLESDEPLATYSMACRLPG
GVNTPQRLWELLERGGRTLSGFPPTDGMWLARLHHPDPPGTSYDCKGFLDDAGF
DAEPFGVSPREAAAMPQOQLLETSMENAGIDIPHSIRGATGFLVAGVAGGE
DTAAADVEGYSVTGVAFAVASGRISYTWMLBGPISVDTGSSSLVALHAAVESLRK
GESSMAVGGAAVATPGVDFPSROBALADGRSARVIRGALPALSOCSEPRVDVAV
LSRBRNGHEVLAIVNGSALNODGANGSASBPRAORIRIOTALSCSEPRVDVAV
EAGGTGTALDPTLANALDVTGRDADBPWLSGSKSTGHTQAAAGVGTGLKVL
ALNGBLPAITLHVEBTPPHVDMSSGGVALLAQNPWRRGRTERRAAVSAGISGTNAH
VIVBERPERHRETTADHGPVPLVVSASTALRAQAAIAELLRPADLISGTNAH
LATTRARHEHRAVASTSEAVRGLREIAAGATADVAVEGTVRVDGRVFLPQO
GQOMAGAGAILSSPPAGKIRACDESNAMPQDMVSDVLRQAPAGLDRVDVOP
VLPAVWVSLAELMRSYGVPRAAVYVHSGOEIAAHYAGALTLEDAAKVYGRRLRS
LSGSGMAVALLGEAAVREKLRPMQDRLPAVNGPRSVVSGPALRPSDCABE
GIRVRDIDVYASHSPQIERVREBELLETTGDIAPRARVTFHSTVESRMDGTLEAB
YWRNRLRETVRFADAVTRLAESGYDAFIEVSPHVVOVVEABEADGEDVAVVGS
LHRDGDLSAFLESMAITAHVSGVDIRMDVALPGAAPPALPTYFORRRVYLOPAPAA
ASDELAYRWSMPTIEKDESGNLDGWLVTYPLISPEMTEMLCEATINANGRALRCVD
TSASRTEMAQVAVAGCTGPRGCVLSLSSDESACRPEVPAGAVGLTLVOLADGATDA
PWCILQGAVRTPADDDLPAPQTTAHGPAQVAGLPGKMGGVVLPSPSVDPAALRL
LVAVLGRGGAEDHLAVRDGRGRVVRASLPOSGSRMTPGTYLVAGLSPVQD
LYVMLADRGAEHLVLAGACPGDOLLAIVEBAGASAVVCAODDALEALGDEBTVAV
HAGTITNFGSISEVABEPEFAETIAAKTALLAVLDEVLDGDAVREYVCSVAITWGA
GMAAVAGSAVLDALABHHRARGRCTSVATPMALPGAUVDTGYABERGLRSLDR
AKRTWERYLAAGVSVAVADVDPVLSBGPATRPTRALRELAGGQABEDSGPT
GEPQRLAGLSPEDEQENLLETVAANAVALVGHESAEINVRARFGLIDSTANAL
RRLASTGTGRPASLVFDPHTVLAOHLRALSDADQAAVAVGAGADESPILATV
GIGCRPGGIGSPEOLMRYLABGANLTGFPADRGMIDIGLYHDPDNDGTSYVDVG
FLTDADPDGPGFGITPRBALAMPQOQLLETMAKWEAVRAGIDPDALGDPGVNG
MNGOSYMOILAGCALGEBVDDYCGLGNSAVLSGRIATTFEGEGALVYDRACSSLVGT
HLMQALRREGCSLALAGVTVMSDFTYTFSTOGKLSBDCCKPSAABDPALE
GVAALVLEPLSRARANGOVLAIRGSANVODASNGLAIPNSOSERYIROLALSG
VPAADVVAEAGHGTGTELDPIEAGALITVYGQDRPRLTSGSKNIGHTQAAAGA
GVYKIVLARMHGMPLPSILADELSPHIDGESGAVETREVPENKIPRPAACVSGF
VSGTNAHVIVEBAPAOEABARTGEPPLTSGRSAAVVAAOARALALHARDPEGL
TDAATLTAIGRARADVRAAVITGDRAGVCAEDALABGPBSADAVAAYLSAPKPYL
PRGCGAQWGMGARDLIESRVPABESRCAEALSPYTDKLTLDVYTGDDGPDPEHVD
VLOPVLPSIMVSLAELMRAGVTPAAVNGHSGEIAAHVAGALSLAAKAVVALSO
VARELDQGGWVSAGSRDELFTVLRKMGRAVAAVNVNGTSVAGPAEDLEFPAE
AEAREMKPRRIAVRASHSPVARIEDRLAABEGITTAVAGSPHLSITVGEVITDSA
MASVWYRNLRPVLPEQAVRGLVEGQDPTVEVSPHVPVLMAVEETALHAGAEVTVCV

PTLRBQSGPHEFTRLNLRHNVHGADLRPAVAGRPALPTYPEFHQFMPRPHRP
ADVSLAGDVGAEHRLPLAADVDPVHGAAPVTGRLSTDQEWLAEHVYGGTTLVGSYL
VDLALAGVDGAPLVEBELVQRPVILVAGAGALLRMSVGA PDSGAPRTIDVHAEDVA
DLADAPWSQHATTTLAGVAGRPDRDQMPREDVARI PLDDHTDGLABQGTBYGSPFO
ALNARKDQSVVAEVSIADEBEGVAFHVPVLLAVAQTTLSLALGEBGGKLPANMT
VTLHAGASISVRVAVPAGADMAALVETPDAGHVAATVSLVRSRTEKKEOEPERG
EGEHLADLWRLAEPSGTGRVVAADSDDAVLRSEBEPDVALVRYRBPEDPRAA
RQGVLMALALVRRMLBOBELPGATVIAIATSGAVTYVDDDSVPRGAAMMGVLRCAQ
ESDRVVLDTDAEPGMLPAVPDNPOLALRGDVPVPRLSPLAPSLATLPAGTORLYP
GDAIDSVPFEPADVQPLRAGEVAVDRATGVNTRDVLLALGMPQKADMGVLAEDL
VVTAVPVDVAPAGDRVGLFGAPAPLAVTDHRLARVPDMSDALAVPATAYT
AHLALHDLGIRAGQSVLIHAAAGVGMAVALARABAGAEVLATAGAKKGTALALG
DDEHILASSETGPARFRRRTGRGVVDVLSVTGLEDSDALADGPFVEMGKTD
LRADGPRGRVAVPFDIGEGDDBELGILERVVGLLAGAGLRLPVSMELTGSAPALQ
HMSGRHVKVLVTPAPYDPOCTVITGCTVGLLLARHLVTHRCVRLILVSRG
ADAPGDELRABEJEDIASABEIACTDRIADBSALLDGLRPLTGVHAGVLAEDL
VYSIDEPAVEQVIRARVADAMNHLBTANTGSLF PVLFPSSAASVLAGPQGVAAANE
SLNALALARTRGLPAKALGWGLMAQASEMTSGIDGR IARTVVALPTERBALFDSA
LRQGEVPEPLSINRSLARRAEVPEVLRKMAKULRAQOARAGPNNVRLAGRSB
SDQVAGLAEVLSHAAVAGVGSADQLPERKAKOIGFOSLAVELRNRLGTATGVRSL
PSTLVDPHTPPLAVAEHLADRLSASPAVDIGRDLERKALEALSMDGHDVGRUL
ESLRRMNSRRADPSTSAISBDASDELFLMDQRFGGGEDLGNSHHHHH"
21742..21788
terminator
/note="T7 terminator"
21809..21826
/note="T7 promoter"
21829..21853
/note="lac operator"
21860..31407
/gene="eryAIII (morphed)"

Query Match 46.8%; Score 281; DB 11; Length 31531;
Best Local Similarity 66.7%; Freq. No. 5.3e-20;
Matches 401; Conservative 0; Mismatches 200; Indels 0; Gaps 0;
Qy 1 GCCTCTAGCGCGCGGCAAGCCTTCCTCGACGCGCTTGCCGCTGATGCGAGCAGT 60
Db 26651 GCCTCTAGCGCGCGTCCGAAACGGTTTCTTGAGTGGTTTCCGACGCGCCGCAATCTGAA 25710
Qy 61 GGGCTCGCGATCACTCGATCGCTCGGAGCTCTGTGGGCGCGGAGAACATGCGCGTACC 120
Db 25711 GCGCGACCCGTCACGATGTCCGATGGGGGTTGTGGCGCGTCAGAACATGCGCGGAT 25770
Qy 121 GAGGGGGGGACTACCTTGCGGCGGCGCGCGCGCCATGACCCCGACGGGGGATC 180
Db 25771 GAAAGCGGTGAGTATGTGCTGACGAGCGCTGCGGCAATGACCAAGATCTGCGGCTG 25830
Qy 181 GAGGACCTCGGACCACTTGACGCGCGGAGAACCCGTGGGTGTGCGTGTGACCTGAC 240
Db 25831 GAAAGACTGCAATATCACTGATATCACTGATCACTCCGTCTCAAGTGTGATATGAC 25890
Qy 241 CGGAGCGGATGCTCGAATCTGTTACCGCGCGCGCGCGCGCGCTCTTTCAGCAATC 300
Db 25891 CGCGCGGTTTGTGAGTTGTTCAAGCGCTGCGCGCTCAAGCTCTTGTGTTGATGAATC 25950
Qy 301 GGTGGGGTCCGCGCGCGGCGCGAGAGAACCGGTCAAGAAATCGGATTTCCGCGCGGCTG 360
Db 25951 GCGGAGTGCACGCGCGGAACTCGCCAGATGAAAGAGGAGCTGCGTGGCGCAGCTCTG 26010
Qy 361 GCGTCGATCCCGGAGGCGCAACGTGACGAGCAATGTGCGCGCGGTGTGCGAGCGGAGT 420
Db 26011 GCGGCACTGTCTACCGCGCAAGCGCGCGGACACTGACACCTGATCGGTGCGGAAATG 26070
Qy 421 GCAGCGGTCTGCGGCGCAAGCGCGCAAGCGGTATCGAGCTGACGCTTCCTTCGTCAC 480
Db 26071 GCAGCGGTCTTGTGTCACGCGCGCAAGTGGCGGATTTGACCGGATGTGTCATTCGCGAT 26130
Qy 481 CTGGGATTTGATCTTCATACCGCGCTGACCTTGGCGGAACCGGCTCGCGCGGTGACCGGG 540
Db 26131 CTGGGGTTTGATCTTCATGACGCGCTGACCTGCGGAACCGCTTCGCAACCGCTGACGGGG 26190
Qy 541 GTCCGGGTGGCAAGCACTCGTCTTGCACCAACCGCAAGTGGAGCGGCTTCACCGGCGAC 600

Db 26191 GTAGGTGAGGCTGCACAGTTGATTGACCATCAACATACAGTACGCGCTGGCGGATCAT 26250
QY 601 T 601
Db 26251 T 26251

RESULT 10
AY623658/c
LOCUS DEFINITION
AY623658 61845 bp DNA linear BCT 16-FEB-2005
Aeromicrobium erythreum putative transcriptional repressor and
putative dehydrogenase/reductase genes, complete cds; erythromycin
biosynthesis gene cluster, complete sequence; putative
oxidoreductase and lipin genes, complete cds; monamine oxidase
gene, partial cds; and unknown gene.
AY623658
AY623658.2 GI:59723038

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Aeromicrobium erythreum
Aeromicrobium erythreum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Propionibacteriaceae; Nocardioidaceae; Aeromicrobium.
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
15557441
2 (bases 1 to 61845)
Brikun,I.A., Reeves,A.R. and Weber,J.M.
Submitted (11-MAY-2004) Fernalogic, Inc., W. Campbell Park Drive,
Chicago, IL 60612, USA
3 (bases 1 to 61845)
Brikun,I.A., Reeves,A.R. and Weber,J.M.
Direct Submission
Submitted (14-FEB-2005) Fernalogic, Inc., W. Campbell Park Drive,
Chicago, IL 60612, USA
Sequence update by submitter
On Feb 14, 2005 this sequence version replaced gi:53794556.
Location/Qualifiers
1. .61845
/organism="Aeromicrobium erythreum"
/mol_type="Genomic DNA"
/db_xref="taxon:2041"
complement(228..869)
/note="ORF13; cecr family"
/codon_start=1
/transl_table=1
/product="putative transcriptional repressor"
/protein_id="AAU93784.2"
/db_xref="GI:59723038"
/translation="MTASAPGESEVSLRAVCAAGVQMTLYHFGNKGQLTDAYVERG
FDLYIAEKSAESTGSDPIQDIRMDAIVAGLNPFGYTLMTGKVRGYSPSAQAP
SELRLBELTRKAHQRLCTTPROAAHVLVNIQVTLQIVLATEEDRLDLSIVAEVGI
AATGSASSPSLQADVRLVLEHANTQRLDLSGSETALLTKMLSLAVGSDG"
925..1746
/note="ORF12; dehydrogenase/reductase fabG family"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase/reductase"
/protein_id="AAU93785.1"
/db_xref="GI:53794558"
/translation="MSLOYSCIADPCSLDIEGYTMSHTLBEKTVVAVAGAKLGG
LISROAAGANVNLATNSASTRADAEETLAAVEAGAGVILTGLDLPVNDALFA
DAESALGADVAVNTVGVLRKPLADTTEADYDAFIDNSKAAFFIIEARHVALE
KVITIVLSLGAFTDGYSTVAGKSPVHFTRAAKAEPAERGI SVTALAPGMDTPPF
YGOETPERVEFHRSGAQMGQLTQIEDISPIVRFLATEGMWITGQITIPANGVYTR"
2172..2729
/note="ORF11"
/codon_start=1
/transl_table=1
/product="unknown"

CDS

gene
CDS
/protein_id="AAU93786.2"
/db_xref="GI:59723040"
/translation="MPSASRRGRRTLLHADCDSPAOPENPACLPMPHPPSSTRRTNP
QSRPASGQRRATPESDSSSSRPTSRVAATSGGVALVGLATRSALIEGHPPTSGAAO
LSTSEIGCFRGAITEQRHEHQAPDAOTPPRPDRPCNKTPTPIYRRPARRAGADDER
GIRRRRTLATLRDAVDPRGPDGLAG"
2920..3813
/gene="eryD1"
2920..3813
/gene="eryC1"
3822..4937
/gene="eryC1"
3822..4937
/gene="eryC1"
/codon_start=1
/transl_table=1
/product="trpD-glucose synthase"
/protein_id="AAU93787.1"
/db_xref="GI:53794560"
/translation="MRGILLAGGSRRLHPITRGISKOLVYVDKPMVYVPLSTMLA
GIRDILITSPHEALFRLLDGSAAGVNLSTAVOPTPGGLAQLIGSSHLCQERA
ALVLGDNITFHGTGQVQLRRHTQTEVAITGRTSNFOAGVVEIGDGRALSLERK
QRPSPYAVPGLFYEPDDVVEHAKTLVPSARGLEITDLRLYLQGRHLVEVLRGT
AMFDGTGFDLNDASNFVRTIEARQGLVGSPEVAVRVGFLSDABILLERHSSST
YGYLRALDREGDADHPYRV"
3822..4937
/gene="eryC1"
3822..4937
/gene="eryC1"
/codon_start=1
/transl_table=1
/product="transaminase"
/protein_id="AAU93788.2"
/db_xref="GI:59723041"
/translation="MTIDVPPFLDIDAAVEIRPOIDAAGVGLAAGRYLLQAGTARE
AEFAACSAHCVVSGCDALBLSLALGVDPQDEVVVAHFTIAMLVSRGAP
VPSPEPDGLVDEVAEVAITPRTAILLVHLYGEADVAALIRVDRGLALVEA
AOGTARGDGAAGVSGSTAAFSFYGKQGLAGDGAAYTADDDARRRLRNFG
SVAKYHVRAGTNSRLDIEQAIVLRLPLIDMLNARBATIAYLRLEASGVRLP
PHDGRSAHMLFVRCRDLTQELTRHGVETLVHPTPVHLSAAYADHGRAGSLP
HAERLARVSLPWPGLPLPAERVADVAVLAR"
complement(4934..5527)
/gene="eryB11"
complement(4934..5527)
/gene="eryB11"
/codon_start=1
/transl_table=1
/product="trpD-4-keco-6-deoxyglucose 3,5-epimerase"
/protein_id="AAU93789.1"
/db_xref="GI:53794562"
/translation="MRAQRVAGGMEFTPVVHDAGVPSPIQEEAFVAAVGHPP
VAQTNHLVSDRGVLRGVHFTASPEQSKTVHCAARALDMLDRTDSTFGTMAVEL
DPRCNALYEPPIGTAHAFMALPEQTINSYLVSTGYDABLIALIDFPDELLPMPADV
PALLSGRDEVAMSPAQARAGLLPRSGPRVABGEVP"
complement(5530..6969)
/gene="eryCV"
complement(5530..6969)
/gene="eryCV"
/codon_start=1
/transl_table=1
/product="deoxyhexose reductase"
/protein_id="AAU93790.1"
/db_xref="GI:53794563"
/translation="METSMMKLVDAVARGVDHTKARDVABRLRSMGSDGVAKAAHLA
ALVSHYAHFPTPLEQORGLDRDSFAALVLEFDRVPELALAVEGPPAKYNSNTI
KPLNEAGALDAVERRRPAFVSVGLYKPTCMFCCHCVRTGARVYADSLTPNEVL
ASVLEVDNDORTTMYSGGLEPLNQLGELITRANRGGLDLMYNSFALTATLS
ROGVNPLGALRTSLIDVLEDEREETTKGA FERVANLORFMHGOEBSRDLIF
NHITLPRGTGLIDVLPFLAEINSSDPRPLDPTVYEDVGRDGLDREBRDLH
GLDFMEYAEKHTGMTIDGYALBSLRGVDAELIRPVTMRPTAHQVAVOYDL
GDVLYBAGPDDLDGARGVYAGVTSDDHSLHVDLRFVESGAEVBPAGDSEYLDGR
DQSVTALNOLDQVDGWRGRLVVAS"
complement(6997..8202)
/gene="eryCIV"
complement(6997..8202)
/gene="eryCIV"
/codon_start=1

gene
CDS

AUTHORS Mochizuki,S., Hiraetsu,K. and Kinashi,H.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2002) Haryasu Kinashi, Hiroshima University.
Department of Molecular Biotechnology, Graduate School of Advanced
Sciences of Matter, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima
739-8530, Japan (E-mail:kinashi@hiroshima-u.ac.jp).
Tel.:81-824-24-7869, Fax:81-824-24-7869)
The nucleotide sequence has been determined by using restriction
fragments and nested deletion fragments of the ordered cosmid
library of pSLA2-L.
pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries
143 ORFs.
Gene prediction was based on the unique codon usage in Streptomyces
(Bibb et al., Gene 30:157-66 (1984)) using the Frameplot program of
Isihkawa and Hotta (FEBS Microbiol Lett 174: 251-253 (1999)) as
implemented at
http://www.nih.go.jp/~junc/cgi-bin/frameplot.pl. Where possible we
chose an initiation codon (atg, gtg, ctg or tgg) which is preceded
by an upstream ribosome binding site sequence (optimally 5-13 bp
before the initiation codon). If this could not be identified we
chose the most upstream initiation codon.
Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes
(three PKS gene clusters for lankacidin, lankamycin and an unknown
type II polyketide, a carotenoid biosynthetic gene cluster, many
regulatory genes and others). The range of each biosynthetic gene
cluster has been deduced by comparison with similar gene clusters
in most cases and may be revised in future.

FEATURES
SOURCE
Location/Qualifiers
1..210614
/organism="Streptomyces rochei"
/mol_type="genomic DNA"
/strain="743AN4"
/db_xref="taxon:1928"
/plasmid="pSLA2-L"
/note="linear plasmid"
1..1992
/note="left terminal inverted repeat, TIR-L; shows 99.4 %
(1981/1992) sequence identity to TIR-R (complement
(208623..210614))"
683..2188
/note="N-terminal sequence is almost identical (435/437)
with that of ORF143 at the right end of pSLA2-L until the
inner end of TIR-L
ORF1 (501 aa)
similar to AL590463 Streptomyces coelicolor putative
helicase, SCPI.136 (879 aa); homology is seen until the
inner end of TIR-L"
/codon_start=1
/transl_table=1
/product="putative helicase"
/protein_id="BAC76459.1"
/db_xref="GI:30698346"
/translacion="MSTTSRTDQRBAAQDEAVDAVVRALIEPVLAPERGLRQVIM
AYGSKTRVARSAEKLAKGVLVLPBLDTQTEAMRAGRTGPMIVGSLRGED
VAPNTDVEELVDMVWPFDVTVFATASGLTLEKRAHKGJLPGMDLIVDAHRT
SGRLGKPAVAVHNDTRIPSLRLVMTATPRLMODDEBAGAGELVASMEDPDGLG
ARCTTLTSLAIDGICAPYOVVCVDTIDTLOAAQLIGVGRSDEVGARLALQTA
LTKASSENPRRTLVFHHMYKEAPAPAGLSDVAKRLAAAPGLVPRITVANNICGHI
KPKHRRVIGFTSGIATDGTVEKGFAGSKVAGSDVTRCDSVWADVGRSHPDL
VQVAFRALRIQPGGKVASLVAVPVLDDGETVADNMLTSPYNGLALLFRQAPLLTGHG
SEGFRAAFPGFEVAPVGVRAQGRGASGAPASRSITPSAMSSISISMRAP"
complement (2315..3595)
/note="ORF2 (426 aa)
similar to AB004736-10 Pseudomonas aeruginosa hypothetical
protein (442 aa)"
/codon_start=1
/transl_table=1
/product="putative secreted protein"
/protein_id="BAC76460.1"
/db_xref="GI:30698347"
/translacion="MRRKPRALLGLPIGLLVAVAPAPAHASADTVVGTAAAPTAA
PAADSTVEDQTLQTVHLVAGDSLRAFRNFEPTSPLTIGEVAAAPAGQAPVAV
DGTDRVVRFGGRPSATLAPGTQWSDVVALPTTAGDLVLSLYLPORTPGSTVSHAA

CDS
ORF3 (328 aa)
similar to Y00459-2 Streptomyces griseus regulatory
protein, Strr (350 aa)"
/codon_start=1
/transl_table=1
/product="Strr-like regulatory protein"
/protein_id="BAC76461.1"
/db_xref="GI:30698348"
/translacion="MTPAALAGNDQETRSIPISLVGSGPRLCGEDAEHVARLADTD
BALPILVERSTRTIDGMHRYLAKXKGTITVELFDGAAASAFLAAYRNHTDL
PLRQBRRAAQRITIAQWPHLSRPAVAGIGAKYALAPRAAGTTPAPORAGSD
GRIRPLDGTGRKKAABLQRPQASVREYARHAGISPATASVDRRLLAGRVEPER
NTGARPAAGTGRATPGGAAPSVSVRIPPEDSPLVRLDRSDLSHKSGRL
LRLQCAVARTLALMAQVPPHCTDLVAELREYADLWAFARVARTDG"
complement (4922..5812)
/note="ORF4 (422 aa), lankacidin biosynthesis protein
similar to AB004625-4 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein B (304 aa)"
/codon_start=1
/transl_table=1
/product="pyrroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76462.1"
/db_xref="GI:30698349"
/translacion="MILGTAAGGGFPQNNCAALCARGRGELPARSGECVAVSGDG
RDWMLNASPIDRITOLLAAPALTPGPBPRTPPAGVLTATBAVHALGLVLRKATGL
TVYAAPVRGALSLELPGRLDRYAPGRADYATAPGFAVAGLTVTYAPVGTAKK
YAAHPDDAPVACVRIEDPATGALVYACLTATWPGFDLTSATCALDGGFFSA
GELGATSSAGAGOSLMGHLPVAGPGSLAALAHRLRRTYTLNNTNPLDPS9AA
HAARVAGVEVLPDGSLEY"
complement (5815..6909)
/note="ORF5 (364 aa), lankacidin biosynthesis protein
similar to A1277117-5 Gluconobacter oxydans putative pqg
protein (359 aa)"
/codon_start=1
/transl_table=1
/product="pyrroloquinoline quinone biosynthesis protein E"
/protein_id="BAC76463.1"
/db_xref="GI:30698350"
/translacion="MADPAGAPACMLIELTRCPHCPYCSNPLEVREBEATCEQ
WTOILQARBLGVOMHFGSGEPRLAPDLVDIVHARLIGAYVVLVSGVGLTBRSH
DIARQYDHYQLSQQDAPPAAGALIAARVHTATLGAAYTAAGLPLTVYIVHGN
IDRTGRVVDLAVDQADRIELANTQYIGWRNRALAPFTAAQALAAAREAVRARTRI
AGGELVYVADVYDDRPKQMDGSGTOLIVTPAGVLPAPAAVATTTLPVSNALRR
PLSEIYASRSFNAYRGTMWRBECRCPSRHADHGGCRCAFQLTGDAATDPAQGL
SPHSLVDAALAEVTDGVPVAFVPRGVPA"
complement (6902..7174)
/note="ORF6 (90 aa), lankacidin biosynthesis protein
similar to AL603642-200 Shiroizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein D (98 aa)"
/codon_start=1
/transl_table=1
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translacion="MTGILPEPTVRLRPGVRLTRDPARGSLALLPERVVVINDTAAAV
LAHCDGTTSLAGIYERLAEYGVASADVEELRLAQRRVVDLHG"
complement (7171..7899)
/note="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AB004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=1
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translacion="MSMSVTRKVAVAPNSEAFRQRLALBSSYWDRRPFHRRMBGLI
DEGLRLMAANRWYORCLPKQDAIYVANCPLPEVRKQMSRIYVHIDADAAGAAEK

CDS

MLTAAVGLRDEVDRLVLAGTEPAVDAYVDFARRRPMLEAASGLTEFSPGL
 ARLRLREHYPMIAEGFEYFARIEVVEPEGRSLDVLARRAVSRBQEQACRALA
 PKRVLAVALDISLDYHNGATRS"
 complement (7926..8030)
 /note="ORF8 (34 aa), lankacidin biosynthesis protein
 similar to AL603642-19 Sinorhizobium meliloti putative
 pyrroloquinoline quinone biosynthesis protein A (31 aa)"
 /codon_start=1
 /transl_table=11
 /product="pyrroloquinoline quinone biosynthesis protein A"
 /protein_id="BAC76466.1"
 /db_xref="GI:30698353"
 /translation="MRTSGKELPAKXAMHRPDVTITDMEVAYFSR"
 8420..10057
 /note="ORF9 (545 aa), possible lankacidin resistance
 protein
 similar to AL133424-19 Streptomyces coelicolor putative
 ABC transporter ATP-binding protein (544 aa)"
 /codon_start=1

CDS

Query Match 40.5%; Score 243.6; DB 1; Length 210614;
 Best Local Similarity 65.1%; Pred. No. 1.5e-16;
 Matches 401; Conservative 0; Mismatches 194; Indels 21; Gaps 2;
 QY 1 GCCTCTACGCGGCGGCAAGCCCTTCTCGACGCTTCGCGCGTGGTGGCGCAGT 60
 Db 60245 GGGCGCTACCGCGGCGGCAAGCCCTTCTCGACGCTTCGCGCGTGGTGGCGCAGT 60186
 QY 61 GGGCTGCGCGTCACTTCGATCGCTGGGATCTGTGGGCGGCGGCAAGCATGGCGGTACC 120
 Db 60185 GGGCTGCGCGTCACTTCGATCGCTGGGATCTGTGGGCGGCGGCAAGCATGGCGGTACC 60126
 QY 121 GAAGGCGGCGCACTACCTGCGCAGCGGCTTCGCGCGCATGAGACCGCGCGGCGATC 180
 Db 60125 GCGGCGCAAGCATCTTCGCGCGCAGCGGCTTCGCGCGCATGAGACCGCGCGGCGATC 60066
 QY 181 GAGAGGCTGCGCAAGCCCTGCGCAGCGGCGGCAAGCGTGGGATCTGTGGGCGGCGGAC 240
 Db 60065 GCGGAGCTGCGCAAGCCCTGCGCAGCGGCGGCAAGCGTGGGATCTGTGGGCGGCGGAC 60006
 QY 241 CGGAGCGGATTCGTGCAACTGTTCAACCGCGCGCGCGCGCGCTCTTTCGACGAACTC 300
 Db 60005 CGGCGCGGATTCGTGCAACTGTTCAACCGCGCGCGCGCGCGCTCTTTCGACGAACTC 59946
 QY 301 GGTGGGATCG 351
 Db 59945 GCGCGGATCG 59886
 QY 352 CGGCGGCTGG-----CGTGAATGCGGAGGCGGAGCGTCAAGCAATGTCGCC 399
 Db 59885 TCCCGGCTGGCGCAAGCTCTGCGCTGCGCGCGCGGAGCGCGCGCAACCTGCTG 59826
 QY 400 CGGCTGTCGAGCGGAGGTGCGAGCGTGTGCGGCGCAAGCGCGCAAGCGTATGAG 459
 Db 59825 CGGCTGTCGAGCGGAGGTGCGAGCGTGTGCGGCGCAAGCGCGCAAGCGTATGAG 59766
 QY 460 CGTGAAGTGTCTTCTGTAAGCTTGGGATTCGATCGATGACCGCGCTGCGCGAAG 519
 Db 59765 CGGGAACCGCGCTTCTGTAAGCTTGGGATTCGATCGATGACCGCGCTGCGCGAAG 59706
 QY 520 CGGCTGCGGCGGCGTGAAGCGGCGGTCCGAGGTGCGCAAGCATGCTTTCGACGAC 579
 Db 59705 CGGATCAACCGGCGGCGGCGGCGGCTGCGGAGGCGACCACTGATCTTCGACGCGCGCAAGT 59646
 QY 580 GTGAGACCGCTCAACG 595
 Db 59645 GTGACCGCGCTGCGCG 59630

RESULT 12
 AX697993 4725 bp DNA linear PAT 02-APR-2003
 LOCUS AX697993
 DEFINITION Sequence 17 from Patent WO03010193.

ACCESSION AX697993
 VERSION AX697993.1 GI:29499042
 KEYWORDS
 SOURCE Micromonospora carbonacea
 ORGANISM Micromonospora carbonacea
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE
 1 Earnet, C.M., Staffa, A. and Yang, X.
 Genes and proteins for the biosynthesis of rosamycin
 Patent: WO 03010193-A 17 06-FEB-2003;
 JOURNAL Biosciences Inc. (CA)
 FEATURES
 source location/Qualifiers
 1..4725
 /organism="Micromonospora carbonacea"
 /mol_type="unassigned DNA"
 /db_xref="taxon:47853"

ORIGIN

Query Match 36.6%; Score 220.2; DB 6; Length 4725;
 Best Local Similarity 61.6%; Pred. No. 1.2e-13;
 Matches 370; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
 QY 3 CTCCTACGCGGCGGCAAGCCCTTCTCGACGCTTCGCGCGTGGTGGCGCAGTGG 62
 Db 3879 CGGCTACGCGGCGGCGGCAAGCCCTTCTCGACGCTTCGCGCGTGGTGGCGCAGTGG 3938
 QY 63 GCTGCGGTCACCTCGATCGCTGGGATCTGTGGGCGGCGGCAAGCATGGCGGTACC 122
 Db 3939 GCTGACCGGCAACTCGATCGCTGGGATCTGTGGGCGGCGGCAAGCATGGCGGTACC 3998
 QY 123 GGGCGCGCACTACCTGCGCAGCGGCTTCGCGCGCATGAGACCGCGAGGCGGATCGA 182
 Db 3999 GCGCGAGGATACCTGAGCGCGCGGCGCTTCGCGCGCATGAGACCGCGAGGCGGATCGA 4058
 QY 183 GGAAGTGTGCGCAAGCCCTGCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
 Db 4059 CGGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4118
 QY 243 GGAAGGATTCGTGCAACTGTTCAACCGCGCGCGCGCGCGCGCTCTTTCGACGAACTC 302
 Db 4119 GAGCTGTTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4178
 QY 303 TGGGATTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 359
 Db 4179 GGAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4238
 QY 360 GGGCTGATGCGGAGGCGGCAAGCTGACGAGCATGTCGCGCGGCTGCTTCGACGAACTC 419
 Db 4239 CGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4298
 QY 420 GGAAGGATTCGTGGGCGGCAAGCGCGCAAGCGTATGAGCGTGCCTTCGCGTGA 479
 Db 4299 CGCGCGCGCTTCGTGGGCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4358
 QY 480 CCGGAGATTCGACTCATGACCGCGCTTCGACTGCGGAAACCGGCTTCGCGGCGGCGG 539
 Db 4359 GATCGGATTCGACTCATGACCGCGCTTCGACTGCGGAAACCGGCTTCGCGGCGGCGG 4418
 QY 540 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 599
 Db 4419 CGTGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4478
 QY 600 C 600
 Db 4479 C 4479

RESULT 13
 AX697977 60196 bp DNA linear PAT 02-APR-2003
 LOCUS AX697977
 DEFINITION Sequence 1 from Patent WO03010193.
 ACCESSION AX697977

VERSION	AX697977.1	GI:29499034
KEYWORDS		
SOURCE	Micromonospora carbonacea	
ORGANISM	Micromonospora carbonacea	
AUTHORS	Farnet, C.M., Staiffe, A. and Yang, X.	
TITLE	Genes and proteins for the biosynthesis of rosamycin	
JOURNAL	Patent: WO 03010193-A 1 06-FEB-2003;	
REFERENCE	Scopia Biosciences Inc. (CA)	
FEATURES	Location/Qualifiers	
SOURCE	1..60196	
ORIGIN	/organism="Micromonospora carbonacea"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:47853"	
Query Match	36.6%; Score 220.2; DB 6; Length 60196;	
Best Local Similarity	61.6%; Pred. No. 4,8e-14;	
Matches 370; Conservative	0; Mismatches 228; Indels 3; Gaps 1;	
Dn	40170 CGCGTACGGGCGGCGCACAAGCGGCCCTTCGACGCGCTCGCGTGCGCGAAGCGGCGCGCGCG	40229
Qy	3 CTCCACGCGGCGGCGCAAAGCGCTTCCTCGACAGCGCTTCGCGCGTGCGTGCGCGCGAGTGG	62
Dn	40170 CGCGTACGGGCGGCGCACAAGCGGCCCTTCGACGCGCTCGCGAAGCGGCGCGCGCGCG	40229
Qy	63 GCTGCGGTGCACCTCGATCGCTCTGGGCTCTGTGGCGCGGCGAGAATGACCGTAGCCGA	122
Dn	40230 GCTGACGCGCACCTCGCGCTCTGGGCGCTCTGTGGGCGCTCGGCGGCGATGCGGCGAGCGA	40289
Qy	123 GGGCGGCGACCTACTCTGCGGACGCGCGGCGCTGCGCGCATGGAACCGGCGCGATGGA	182
Dn	40290 CGCGGAGAGTAACCTGAGCGCGCGCGCGCTGCGCGCATGCTCCGAGGTGGCGTGGAA	40349
Qy	183 GGAAGTGGGAGCAACCCTGAGCGCGCGGCGAACCGGTGGTGTGCGATGTGACCTGAGACG	242
Dn	40350 CGCCCTCTGGCGCGCGCTCGAGACCGGAGACGACCTTCTGTGCGCGTGGCGGACGTGAC	40409
Qy	243 GGAAGCGGTTCGTGAACTGTTCACTGCGCGCGCGCGCGCGCGCGCGCTCTTCGAGCAACTCG	302
Dn	40410 GACGCTGTTCACGAGCGCGGATTCACTCGGATTCCGAGCCGACCGCGCTGCTCGGACCTCCC	40469
Qy	303 TGGGGTCCGCGCGCGGCGCGGAGAGACCGGTCT--AGGATCCGANTCTGCGCGCGCGGT	359
Dn	40470 GGAAGCGCGCGCGAGCGCTGAGCGGACGCGCGAACCGCGCGGCTTCGACCTTGCAGCGCTGGA	40529
Qy	360 GGCCTGATGCGGAGCGCGGACGTAACGTAACGATGTCGCGCGGTGATCCGAGCGAGCG	419
Dn	40530 CGCGCGCGGAGCGCGCGGAGGACGCGCGCGCGCGCGCTCTGACCTGTGATGCGCGGAGGT	40589
Qy	420 GGCAGCGGTGCTGGGCGACGCGGACGCGGATGATGAGCGGTGACGTGCGCTTCCGTGA	479
Dn	40590 CGCGCGCGGTCTCGGCGACCGCGGCGCGGAGCATGTCGAGCGCGCGGACGCGCGGTCCGGGA	40649
Qy	480 CCTGGGATTCGACTCATGACCGCGGTGATGACCTTGGGAAACCGGTGCGCGCGGTGACCG	539
Dn	40650 GATCGGATTCGACTCTGCTGACCGCGGTGACCTTGGGAAACCGGTGCGCGCGGTGACCG	40709
Qy	540 GGTTCGGGCGGCGGACGACCATGCTTTCGACCAACCGGACGATGGAACCGCTTCACCGGCG	599
Dn	40710 CGTGGCGGTGCTGCGGACCGCTTGTCTTGACCAACCGGACGCGGCGGTTCGCGGAGCA	40769
Qy	600 C 600	
Dn	40770 C 40770	
RESULT 14		
LOCUS	AB070940	104326 bp DNA linear BCT 22-JAN-2002
DEFINITION	Streptomyces avermitilis oligomycin biosynthetic gene cluster.	
ACCESSION	AB070940	GI:15823967
VERSION	AB070940.1	GI:15823967

KEYWORDS	Streptomyces avermitilis
SOURCE	Streptomyces avermitilis
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE	1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinoue, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osomoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
TITLE	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
JOURNAL	11572948
PIBMED	2 (bases 1 to 104326)
REFERENCE	Ikeda, H.
AUTHORS	Direct Submission
TITLE	Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
JOURNAL	(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)
FEATURES	Location/Qualifiers
source	1..104326
gene	/organism="Streptomyces avermitilis"
gene	/mol_type="genomic DNA"
gene	/db_xref="taxon:33903"
gene	complement(1..1134)
gene	/gene="mcma"
gene	complement(1..1134)
gene	/gene="mcma"
gene	/note="methylmalonyl-CoA mutase"
gene	/codon_start=1
gene	/transl_table=1
gene	/protein_id="BAB69185.1"
gene	/db_xref="GI:15823968"
gene	/translation="MTGSGTADANARFROLVGSGTRAVPVFDLPTFAGIDSDAPAA RGVGRGCAVDSIDDMRWVIFGGRLRVRVPSLAVDPAAPLVLTVLVAREHGAPAG RLAGVAGDVFKETPLARGPGFPFRLPSRLAMNVPACQALPMQCLSGGHLIAGT GADPLAEVAFVAGGIEYLRAVTAAGLDVRAVRLDCLVGTTGSENKRLPATRR VMTPLANERFGSGRTAPLTVRTPTDGLATACBEDIEAGATLAVDRTERLGGTLTVL BHGLAGGLGRPHGDRPATPRPFPDQSLDLQAROTERLKLAMRVQPAVDBALRL LRKAQGDVNLVYMRALALARAATVGCDAALRTWGSPSAT"
gene	complement(1111..1886)
gene	/gene="pcta"
gene	complement(1131..1886)
gene	/gene="pcta"
gene	/codon_start=1
gene	/transl_table=1
gene	/product="phosphopantetheinyl transferase"
gene	/protein_id="BAB69186.1"
gene	/db_xref="GI:15823969"
gene	/translation="MRAHVPADVPVSLHPREALVODAVESRRRSPFARHCCARR LASLGGPPLATPGFQKRWPAVGCAGITACAGRAAVARAGIARVAGIDAEEDR LPGVLAEVALPAGFOYKHTAHHPGVWQKLFSAGSEYKAFPAVGGPLGFESE LIDDPVSTFRVRLPTALRAARELDLHGRLVIRGILTVAVVNDGPRDTPHA SPYDKTSAPLGRSADRPGRPTDRATDRPTA"
gene	complement(1910..2875)
gene	/gene="slmk4 homolog"
gene	complement(1910..2875)
gene	/gene="slmk4 homolog"
gene	/codon_start=1
gene	/transl_table=1
gene	/product="Slmk4 homolog"
gene	/protein_id="BAB69187.1"
gene	/db_xref="GI:15823970"
gene	/translation="MSCRDHGSTAALGGPSHGRLAVSDIHVRHCONRDIVDILH GAEMVLVAGDVASVADITWALSILDRPAXIWAQNHETLWAPDPRTERER YRHLIDCRALGVVTPEDPVVMHGAGQVAAPAFLLYDSFLPGAMSHAKRL DDAGVICTDEFLHDPDPPYSRBAKCAVRVATRALDLPLDPTVLVNHSPVLR RLRHPERFALMGCTATADMLRYAAAVVGHILHPTICIDDPHNEVSLGYPRR QYGRGRDPVQILTRAPARAAPRVVPAVRPGHRTSPADRRP"
gene	3213..3971


```

Db      14764 CCGAAGGCGCGCGCGCGCGCAAGACGATCCGGCCCGGAGAGTCCGAACCTGGCCGCC 14705
Qy      355 CCGCTGGCGCTCGATGCCGAGAGGCCGAACGTCAAGACATGTCCGCCCGGCTGGTCCGAGCC 414
Db      14704 AGGCTGGCCCGGCTCGACCGAAGCGGGAACGCGAACTCCTCGACTGGTCAAGGACC 14645
Qy      415 GAGGTGGCAAGCGGTGCTGGGCGCAAGCGACCGCGATCGAGCGTGAACGTGCGCTTC 474
Db      14644 GCCGCCGCCCTCGGCTGGGCGCAACCGGCGCGCGCAATCACCCGTCGAAAGCCCTTC 14585
Qy      475 CGTGACCTGGGATTCGATCCATGACCGCGTCCGACCTGCGGAACCGGCTCGCGCGGTG 534
Db      14584 AAGGACCTGGGCTTCGACTCCCTGACCGCGCTCGACCTGGGCAACCGGCTCACCGCCG 14525
Qy      535 ACCGGGGTCCGGGTGGCCAGACCAATGTTCTTCGACCAACCGAC 578
Db      14524 ACCGGCGTACGTCTGCCCGCACGCTGCTTCGACCAACCCAC 14481

```

Search completed: April 7, 2006, 16:57:38
 Job time : 2908.81 secs

This Page Blank (uspto)